Journal of Medical Internet Research

Journal Impact Factor (JIF) (2022): 7.4
Volume 24 (2022), Issue 1 ISSN 1438-8871 Editor in Chief: Gunther Eysenbach, MD, MPH

Contents

Reviews

Viewpoints

Original Papers

Letter to the Editor

Corrigenda and Addenda

Review

Accuracy and Acceptability of Wrist-Wearable Activity-Tracking Devices: Systematic Review of the Literature

Federico Germini[1,2], MSc, MD; Noella Noronha[1,3], MSc; Victoria Borg Debono[1], MSc, PhD; Binu Abraham Philip[1], MSc; Drashti Pete[1], MPH; Tamara Navarro[1], MLIS, MEd; Arun Keepanasseril[1,2], MSc, MDS; Sameer Parpia[1,4], PhD; Kerstin de Wit[1,5], MBChB, MSc, MD; Alfonso Iorio[1,2], MD, PhD

[1]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

[2]Department of Medicine, McMaster University, Hamilton, ON, Canada

[3]School of Interdisciplinary Sciences, McMaster University, Hamilton, ON, Canada

[4]Department of Oncology, McMaster University, Hamilton, ON, Canada

[5]Department of Emergency Medicine, Queen's University, Kingston, ON, Canada

Corresponding Author:
Noella Noronha, MSc
Department of Health Research Methods, Evidence, and Impact
McMaster University
1280 Main Street West
Hamilton, ON, L8S 4L8
Canada
Phone: 1 9055259140 ext 22069
Email: noronn@mcmaster.ca

Abstract

Background: Numerous wrist-wearable devices to measure physical activity are currently available, but there is a need to unify the evidence on how they compare in terms of acceptability and accuracy.

Objective: The aim of this study is to perform a systematic review of the literature to assess the accuracy and acceptability (willingness to use the device for the task it is designed to support) of wrist-wearable activity trackers.

Methods: We searched MEDLINE, Embase, the Cochrane Central Register of Controlled Trials, and SPORTDiscus for studies measuring physical activity in the general population using wrist-wearable activity trackers. We screened articles for inclusion and, for the included studies, reported data on the studies' setting and population, outcome measured, and risk of bias.

Results: A total of 65 articles were included in our review. Accuracy was assessed for 14 different outcomes, which can be classified in the following categories: count of specific activities (including step counts), time spent being active, intensity of physical activity (including energy expenditure), heart rate, distance, and speed. Substantial clinical heterogeneity did not allow us to perform a meta-analysis of the results. The outcomes assessed most frequently were step counts, heart rate, and energy expenditure. For step counts, the Fitbit Charge (or the Fitbit Charge HR) had a mean absolute percentage error (MAPE) <25% across 20 studies. For heart rate, the Apple Watch had a MAPE <10% in 2 studies. For energy expenditure, the MAPE was >30% for all the brands, showing poor accuracy across devices. Acceptability was most frequently measured through data availability and wearing time. Data availability was ≥75% for the Fitbit Charge HR, Fitbit Flex 2, and Garmin Vivofit. The wearing time was 89% for both the GENEActiv and Nike FuelBand.

Conclusions: The Fitbit Charge and Fitbit Charge HR were consistently shown to have a good accuracy for step counts and the Apple Watch for measuring heart rate. None of the tested devices proved to be accurate in measuring energy expenditure. Efforts should be made to reduce the heterogeneity among studies.

(J Med Internet Res 2022;24(1):e30791) doi:10.2196/30791

KEYWORDS

diagnosis; measurement; wrist-wearable devices; mobile phone

Introduction

Background

Tracking, measuring, and documenting one's physical activity can be a way of monitoring and encouraging oneself to participate in daily physical activity; increased activity is thought to translate into important positive health outcomes, both physical and mental [1]. In the past, most physical activity tracking was done manually by oneself or an external assessor, through records, logbooks, or using questionnaires. These are indirect methods to quantify physical activity, meaning that they do not measure movement directly as it occurs [2]. The main disadvantages of such methods are the administrative burden on either the self-assessor or the external assessor and the potential imprecision because of recall bias [2,3]. Direct methods to assess physical activity [2], such as accelerometers or pedometers that digitally record movement, are preferred because they eliminate recall bias and are convenient. This process of activity tracking has become automated, accessible, and digitized with wearable tracking technology such as wristband sensors and smartwatches that can be linked to computer apps on other devices such as smartphones, tablets, and PCs. When data are uploaded to these devices, users can review their physical activity log and potentially use this feedback to make behavior changes with regard to physical activity.

The ideal device should be acceptable to the end user, affordable, easy to use, and accurate in measuring physical activity. Accuracy can be defined as the closeness of the measured value to the actual value. Accuracy can be calculated using measures of agreement, sensitivity and specificity, receiver operating characteristic curves, or absolute and percentage differences [4]. Agreement can be defined as "the degree of concordance between two or more sets of measurements" [5]. It can be measured as percentage agreement, that is, the percentage of cases in which 2 methods of measurements of the same variable lead to classification in the same category. Another example of methods of calculating agreement is the κ statistic, which measures agreement beyond chance [6]. Sensitivity and specificity are the true positive and true negative proportions, respectively. These proportions are calculated using the measurement method that we are evaluating as the index test and another method, known to be accurate, as the reference standard [4]. Receiver operating characteristic curves are obtained plotting the sensitivity versus the complement of specificity and can be used to find optimal cutoff points for the index test. Absolute and percentage differences are used to determine how far the index test measurement is from the reference standard or their average [4]. Acceptability can be widely defined as "the demonstrable willingness within a user group to employ information technology for the task it is designed to support" [7]. It can be assessed qualitatively (eg, through questionnaires or interviews) or quantitatively (eg, percentage of the time during which the device is worn or the data are available or measured using ad hoc scales). On the basis of a 2019 review, acceptability or acceptance of wrist-wearable activity-tracking devices is dependent on the type of user and context of use [8]. This same review indicates that research on accuracy has not kept up with the plethora of wearable physical activity–tracking devices in the market [8]. This may be because of the rapidly changing landscape as companies continue to upgrade models with different technical specifications and features. The purpose of this systematic review is to assess the acceptability and accuracy of these wrist-wearable activity-tracking devices through a focused in-depth review of primary studies assessing these 2 characteristics.

Objectives

The first objective of this systematic review is to assess the accuracy of wrist-wearable activity-tracking devices for measuring physical activity.

The second objective is to assess the acceptability of wrist-wearable activity-tracking devices for measuring physical activity.

Methods

The methods used for this systematic review have been registered in the PROSPERO database (CRD42019137420).

Search Strategy

The databases searched were MEDLINE, Embase, the Cochrane Central Register of Controlled Trials, and SPORTDiscus from inception to May 28, 2019. Search strategies were developed to retrieve content on wearable activity trackers and on their accuracy and reproducibility of results. We used search terms, including *Wearable device* and *Fitness tracker*, to identity studies on the use of a consumer-based wearable activity tracker, whereas terms such as *data accuracy* and *reproducibility of results* were included to bring in content focused on activity tracker validation. The search strategy is available on the web in the PROSPERO record. A snowball search was conducted by checking the references of relevant studies and systematic reviews on this topic that were identified in our original search.

Selection of Studies

For the acceptability objective, the population was the general population, without sex or age restrictions. The intervention was the use of a wrist-wearable activity tracker. The outcome was any quantitative measure of acceptability, including wearing time, data availability, and questionnaires to assess acceptability.

For the accuracy objective, the population was again the general population, the index test had to be a wrist-wearable activity tracker, and the reference standard could be another device or any method used to measure physical activity, including questionnaires and direct observation. The outcome could be any measure of physical activity, including but not limited to step count, heart rate, distance, speed, activity count, activity time, and intensity of physical activity.

For both objectives, this review examined both research-grade devices (activity trackers available only for research purposes) and commercial devices (those available to the general public). The included studies were limited to the community-based everyday-life setting. Laboratory tests such as research studies were included as long as everyday settings were reproduced, thereby excluding patients who were institutionalized and those

who were hospitalized. We set no restrictions on the length of observation for the original studies.

The exclusion criteria included the following: device not worn on the wrist, studies measuring sleep, and studies on patients who were institutionalized or hospitalized.

All studies reporting primary data were considered for inclusion, with the exception of case reports and case series.

Using the aforementioned inclusion and exclusion criteria and a piloted form, we initially screened for inclusion from the titles and abstracts of the retrieved articles, using the web-based software Rayyan (Rayyan Systems Inc) [9]. Subsequently, we screened the full texts of the studies identified as potentially eligible from the title and abstract screening for selection.

Data Extraction and Risk-of-Bias Assessment

Data were extracted to an Excel (Microsoft Corp) file. The data extraction form was based on a previous publication on the same topic [8] and adapted to the needs of this review. The following data were extracted: general study information: first author's name, publication year, type of study (prospective vs retrospective and observational vs interventional), duration of follow-up (in days), and setting (laboratory vs field); characteristics of the population: number of participants, underlying health condition (eg, healthy participants, people with severe obesity, and chronic joint pain), gender, and age distribution (mean and SD or median and minimum–maximum or first and third quartiles); measures of accuracy: step count, distance, speed, heart rate, activity count, time spent being active, and intensity of physical activity; and acceptability of the device, including but not limited to data availability, wearing time, ease of use. The risk of bias was assessed using the Quality Assessment of Diagnostic Accuracy Studies, version 2, tool [10]. This tool guides the assessment of the risk of bias in diagnostic accuracy studies in 4 domains: patient selection, index test, reference standard, and flow and timing. We rated the risk of bias in each domain as *High*, *Probably high*, *Probably low*, and *Low*. When necessary, the study authors were contacted for additional information.

Throughout title and abstract and full-text screening and the data extraction, each step was performed in duplicate with 2 reviewers (NN and BAP) deciding independently on inclusion or exclusion and, if needed, later having a discussion with another author to make a final decision. Disagreements were solved through discussion and, when needed, with the intervention of a third reviewer (FG, VBD, or DP). The reviewers were trained with calibration exercises until an adequate performance was achieved for each of these steps.

Diagnostic Accuracy Measures

When available, we extracted the mean absolute percentage error (MAPE) or the mean percentage error. When these were not available, we extracted other measures in the following order of priority: mean difference, mean bias (Bland–Altman), accuracy determined through intraclass correlation coefficient, and correlation coefficient (Pearson or Spearman). When the outcome was dichotomized and sensitivity and specificity were calculated, we reported on these values. When available, we reported measures of variability or 95% CIs for all the aforementioned measures. The formulas used for calculating the MAPE, mean percentage error, mean difference, and mean bias are reported in Multimedia Appendix 1 [11-75].

Synthesis of Results

Because of the significant heterogeneity observed in the studies' populations, settings, devices assessed, reference standards, outcomes assessed, and the outcome measures reported, we decided not to perform a quantitative synthesis and have provided a narrative synthesis of the results for both the objectives. For the accuracy objective, given the high number of studies retrieved, we summarized results only for devices that were included in at least two studies reporting the same outcome. All the remaining results are reported in Multimedia Appendix 1.

Ethics Approval

This systematic review was based on published data and therefore did not require a submission to a research ethics board.

Availability of Data and Materials

Most of the data that support the findings of this study are available in Multimedia Appendix 1. A guide on how to use the database provided in Multimedia Appendix 1 can be found in Multimedia Appendix 2. The full data set can be made available upon reasonable request.

Code Availability

This is not applicable to this systematic review because no quantitative data synthesis was performed.

Results

Overview

The search identified 1633 records (1614, 98.84%, after the removal of duplicates). The study flow diagram is presented in Figure 1. After screening the full texts of 398 articles, 65 (16.3%) were included in the systematic review. The characteristics of the included studies are summarized in Table 1 and Multimedia Appendix 3 [11-67] for the accuracy objective and Table 2 for the acceptability objective. All the included studies were single-center studies, with a prospective, observational design. The complete results for accuracy and acceptability have been reported in Multimedia Appendix 1.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 study flow diagram.

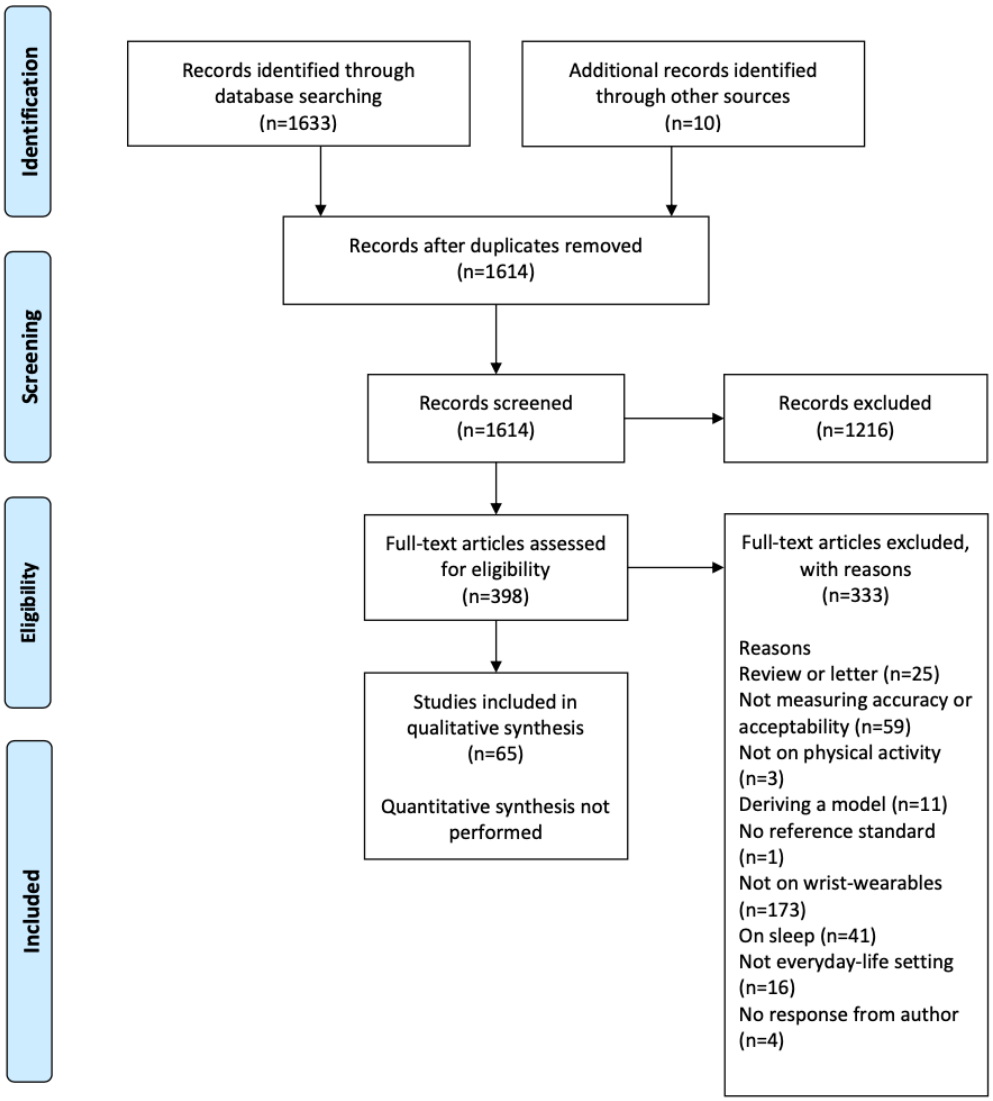


Table 1. Characteristics of the studies reporting on accuracy (N=57).

First author, year	Setting	FUP[a] time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome	Device brand and model
Alharbi [11], 2016	Laboratory	<1	48	66 (7)	48	Patients undergoing cardiac rehabilitation	Step count and MVPA[b]	Fitbit Flex
Alsubheen [12], 2016	Field	5	13	40 (12)	38	Healthy	Step count and energy expenditure	Garmin Vivofit
An [13], 2017	Laboratory and field	1	35	31 (12)	51	Healthy	Step count	Fitbit Flex, Garmin Vivofit, Polar Loop, Basis B1 Band, Misfit Shine, Jawbone UP24, and Nike FuelBand SE
An [14], 2017	Field	<1	62	24 (5)	40	Healthy	Active time	ActiGraph GT3X
Blondeel [15], 2018	Field	14	8	65 (8)	25	COPD[c]	Step count	Fitbit Alta
Boeselt [16], 2016	Field	3	20	66 (7)	15	COPD	Step count, energy expenditure, and MVPA	Polar A300
Bruder [17], 2018	Laboratory and field	395	32	__[d]	—	Rehabilitation after radial fracture	Activity count	ActivPAL
Bulathsinghala [18], 2014	Laboratory	1	20	70 (10)	—	COPD	Physical activity intensity	ActiGraph GT3X+
Burton [19], 2018	Laboratory	<1	31	74 (6)	65	Healthy older adults	Step count	Fitbit Flex and Fitbit Charge HR
Choi [20], 2010	Laboratory	<1	76	13 (2)	62	Healthy	Energy expenditure	ActiGraph GT1M
Chow [21], 2017	Laboratory	1.5	31	24 (5)	39	Healthy	Step count	ActiGraph wGT3xBT-BT, Fitbit Flex, Fitbit Charge HR, and Jawbone UP24
Chowdhury [22], 2017	Laboratory and field	2	30	27 (6)	50	Healthy	Energy expenditure	Microsoft Band; Apple Watch, series not specified; Jawbone Up24; and Fitbit Charge
Cohen [23], 2010	Laboratory and field	3	57	70 (10)	—	COPD	Speed	ActiGraph Mini Motionlogger
Compagnat [24], 2018	Laboratory	<1	46	65 (13)	—	Stroke	Energy expenditure	ActiGraph GT3X+
Dondzila [25], 2018	Laboratory and field	3-62	40	22 (2)	58	Healthy	Step count, energy expenditure, and heart rate	Fitbit Charge HR and Mio Fuse
Dooley [26], 2017	Laboratory	1	62	23 (4)	58	Healthy	Heart rate and energy expenditure	Apple Watch, series not specified; Fitbit Charge HR; and Garmin Forerunner 225
Durkalec-Michalski [27], 2013	Laboratory	2	20	26 (5)	55	Healthy	Energy expenditure	ActiGraph GT1M
Falgoust [28], 2018	Laboratory	1	30	—	—	Healthy	Step count	Fitbit Charge HR, Fitbit Surge, and Garmin Vivoactive HR

First author, year	Setting	FUP[a] time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome	Device brand and model
Ferguson [29], 2015	Field	2	21	33 (10)	52	Healthy	Step count, MVPA, and energy expenditure	Nike FuelBand, Misfit Shine, and Jawbone UP
Gaz [30], 2018	Laboratory and field	<1	32	36 (8)	69	Healthy	Step count and distance	Fitbit Charge HR; Apple Watch, series not specified; Garmin Vivofit 2; and Jawbone UP2
Gillinov [31], 2017	Laboratory	<1	50	38 (12)	54	Healthy	Heart rate	Garmin Forerunner 235; TomTom Spark; Apple Watch, series not specified; and Fitbit Blaze
Gironda [32], 2007	Laboratory	<1	3	43[e]	31	Pain syndromes	Activity count	Actiwatch Score
Hargens [33], 2017	Laboratory and field	7	21	31[e]	68	Healthy	MVPA, energy expenditure, and step count	Fitbit Charge
Hernandez-Vicente [34], 2016	Field	7	18	21 (1)	50	Healthy	Energy expenditure, vigorous active time, active time, and step count	Polar V800
Huang [35], 2016	Laboratory	1	40	24 (3)	25	Healthy	Step count and distance	Jawbone UP24, Garmin Vivofit, Fitbit Flex, and Nike FuelBand
Imboden [36], 2018	Laboratory	<1	30	49 (19)	50	Healthy	Energy expenditure, MVPA, and step count	Fitbit Flex, Jawbone Up24, and Fitbit Flex
Jo [37], 2016	Laboratory	<1	24	25[e]	50	Healthy	Heart rate	Basis Peak K and Fitbit Charge
Jones [38], 2018	Laboratory	118	30	33[e]	—	Healthy	Step count	Fitbit Flex
Kaewkannate [39], 2016	Laboratory	<1	7	31 (0)	14	Healthy	Step count	Fitbit Flex, Jawbone UP24, Withings Pulse, and Misfit Shine
Lamont [40], 2018	Laboratory	<1	33	67 (8)	64	Parkinson disease	Step count	Garmin Vivosmart HR and Fitbit Charge HR
Lauritzen [41], 2013	Laboratory and field	<1	18	—	56	Older adults	Step count	Fitbit Ultra
Lawinger [42], 2015	Laboratory	<1	30	26 (6)	70	Healthy	Activity count	ActiGraph GT3X+
Lemmens [43], 2018	Laboratory	<1	40	31 (5)	100	Parkinson disease	Energy expenditure	Philips optical heart rate monitor
Magistro [44], 2018	Laboratory	<1	40	74 (7)	60	Healthy	Step count	ADAMO Care Watch
Mandigout [45], 2017	Laboratory	<1	24	68 (14)	60	Stroke	Energy expenditure	Actical and ActiGraph GTX
Manning [46], 2016	Laboratory	<1	9	15 (1)	—	Severe obesity	Step count	Fitbit One, Fitbit Flex, Fitbit Zip, ActiGraph GT3x+, and Jawbone UP

First author, year	Setting	FUP[a] time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome	Device brand and model
Montoye [47], 2017	Laboratory	<1	30	24 (1)	47	Healthy	Step count, energy expenditure, and heart rate	Fitbit Charge HR
Powierza [48], 2017	Field	<1	22	22 (2)	55	Healthy	Heart rate	Fitbit Charge
Price [49], 2017	Laboratory	<1	14	23[e]	21	Healthy	Energy expenditure	Fitbit One, Garmin Vivofit, and Jawbone UP
Redenius [50], 2019	Laboratory	4	65	42 (12)	72	Healthy	MVPA	Fitbit Flex
Reid [51], 2017	Field	4	22	21 (2)	100	Healthy	MVPA and step count	Fitbit Flex
Roos [52], 2017	Laboratory	2	20	24 (2)	40	Runners	Energy expenditure	Suunto Ambit, Garmin Forerunner 920XT, and Polar V800
Schaffer [53], 2017	Laboratory	<1	24	54 (13)	42	Stroke	Step count	Garmin Vivofit
Scott [54], 2017	Field	7	89	—	54	Healthy	Daily mean activity and MVPA	GENEActiv
Semanik [55], 2020	Laboratory	7	35	52[e]	69	Chronic joint pain	MVPA	Fitbit Flex
Sirard [56], 2017	Laboratory and field	7	14	9 (2)	50	Healthy	Energy expenditure, MVPA, and step count	Movband and Sqord
St-Laurent [57], 2018	Laboratory	7	16	33 (4)	100	Pregnant	Step count and MVPA	Fitbit Flex
Stackpool [58], 2013	Laboratory	<1	20	22 (1)	50	Healthy	Step count and energy expenditure	Jawbone UP, Nike FuelBand, Fitbit Ultra, and Adidas miCoach
Stiles [59], 2013	Laboratory	1	10[e]	39 (6)	100	Healthy premenopausal women	Loading rate (BW^f/s)	GENEActiv and ActiGraph GT3X+
Støve [60], 2019	Laboratory	<1	29	29 (9)	41	Healthy	Heart rate	Garmin Forerunner
Tam [61], 2018	Laboratory	<1	30	32 (9)	50	Healthy	Step count	Fitbit Charge HR and Xiaomi Mi Band 2
Thomson [62], 2019	Laboratory	<1	30	24 (3)	50	Healthy	Heart rate	Apple Watch, series not specified; and Fitbit Charge HR2
Wahl [63], 2017	Laboratory	<1	20	25 (3)	50	Healthy	Step count, energy expenditure, and distance	Polar Loop, Beurer AS80, Fitbit Charge HR, Fitbit Charge, Bodymedia Sensewear, Garmin Vivofit, Garmin Vivosmart, Garmin Vivoactive, Garmin Forerunner 920XT, Xiaomi Mi Band, and Withings Pulse
Wallen [64], 2016	Laboratory	<1	22	24 (6)	50	Healthy	Heart rate, energy expenditure, and step count	Apple Watch, series not specified; Samsung Gear S; Mio Alpha; and Fitbit Charge

First author, year	Setting	FUP[a] time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome	Device brand and model
Wang [65], 2017	Laboratory	<1	9	22 (1)	44	Healthy	Step count	Huawei B1, Xiaomi Mi Band, Fitbit Charge, Polar Loop, Garmin Vivofit 2, Misfit Shine, and Jawbone UP
Woodman, 2017 [66]	Laboratory	<1	28	25 (4)	29	Healthy	Energy expenditure	Garmin Vivofit, Withings Pulse, and Basis Peak
Zhang [67], 2012	Laboratory	1	60	49 (7)	62	Healthy	Activity classification (sedentary, household, walking, and running)	GENEActiv

[a]FUP: follow-up.

[b]MVPA: moderate- to vigorous-intensity physical activity.

[c]COPD: chronic obstructive pulmonary disease.

[d]Not available.

[e]SD not reported.

[f]BW: body weight.

Table 2. Characteristics of the studies reporting on acceptability (N=11).

First author, year	Setting	FUP[a] time, days	Sample, n	Age (years), mean (SD)	Female, %	Underlying health condition	Outcome assessed	Device brand and model
Boeselt [16], 2016	Laboratory and field	7	20	66 (7)	15	COPD[b]	Ease of use and other characteristics	Polar A300
Deka [68], 2018	Field	5	46	65 (12)	67	CHF[c]	Data availability	Fitbit Charge HR
Farina [69], 2019	Field	2	26; 26	80 (6); 76 (6)	39; 73	Dementia; caregivers of patients with dementia	Wearing time	GENEActiv
Fisher [70], 2016	Field	7	34	69[d]	—[e]	Parkinson disease	Ease of use and other characteristics	AX3 data logger
Kaewkannate [39], 2016	Field	<1	7	31 (0)	14	Healthy	Ease of use and other characteristics	Fitbit Flex, Jawbone UP24, Withings Pulse, and Misfit Shine
Lahti [71], 2017	Laboratory	120	40	—	—	Schizophrenia	Data availability	Garmin Vivofit
Marcoux [72], 2019	Field	46	20	73 (7)	20	Idiopathic pulmonary fibrosis	Data availability	Fitbit Flex 2
Naslund [73], 2015	Field	80-133	5	48 (9)	90	Serious mental illness	Wearing time	Nike FuelBand
Speier [74], 2018	Laboratory	90	186	—	—	Coronary artery disease	Wearing time	Fitbit Charge HR2
St-Laurent [57], 2018	Laboratory	1	16	33 (4)	100	Pregnant	Ease of use and other characteristics	Fitbit Flex
Rowlands [75], 2018	Field	425	1724	13 (1)	100	Healthy	Data availability	GENEActiv

[a]FUP: follow-up.

[b]COPD: chronic obstructive pulmonary disease.

[c]CHF: congestive heart failure.

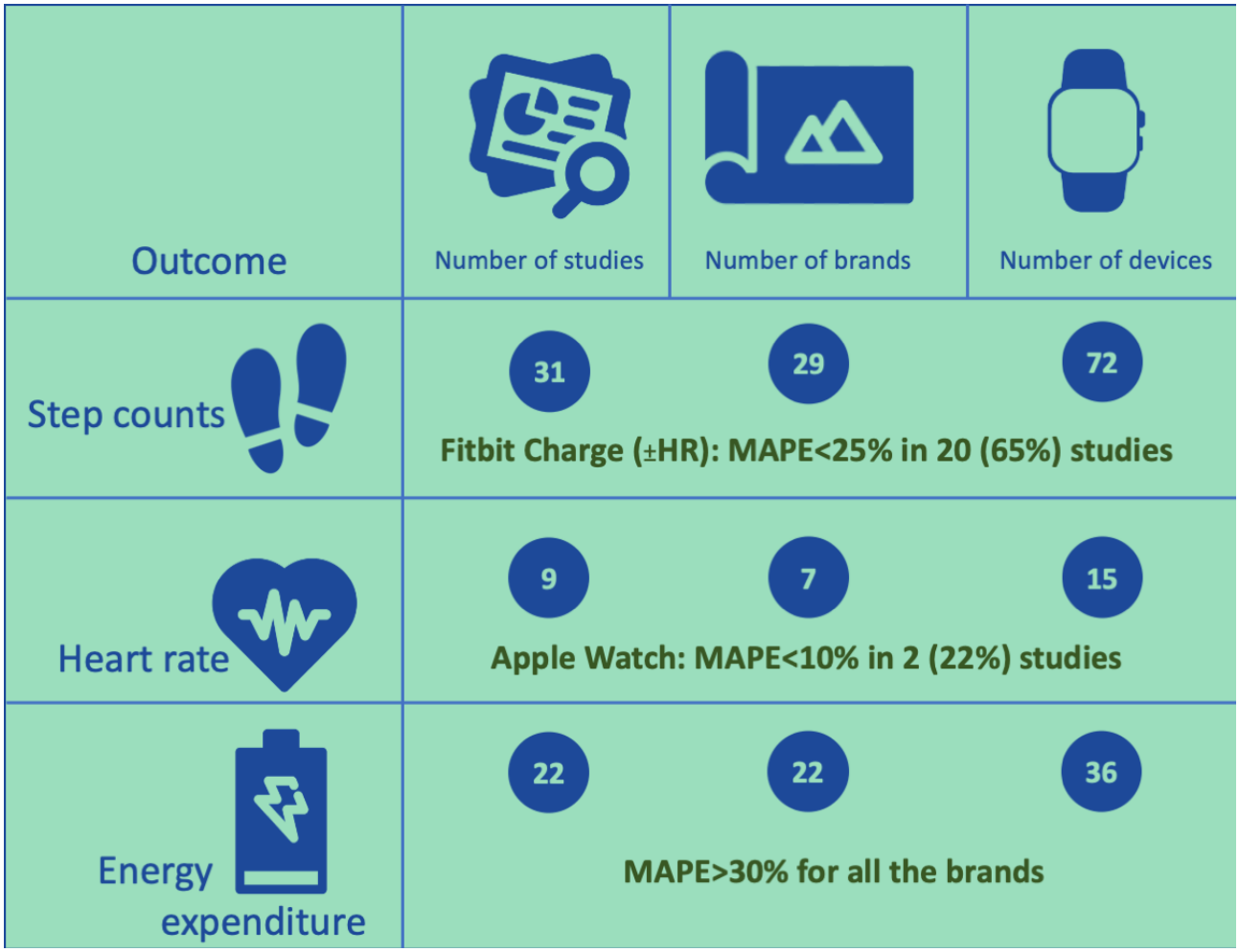
[d]SD not reported.

[e]Not available.

Accuracy

The accuracy of wrist-wearable activity trackers was assessed in 57 studies on 72 devices from 29 brands. Step count, heart rate, and energy expenditure (EE) were the most commonly assessed outcomes in the appraised literature. The results of these outcomes are summarized in Figure 2 (icons by Nikhil Bapna, Yoyon Pujiyono, Chintuza, Gregor Cresnar, Andrejs Kirma, and Yigit Pinarbasi from the Noun Project [76]), in which we have highlighted the standout device for the most frequently reported outcomes.

Figure 2. Summary of the results for the main accuracy outcomes. MAPE: mean absolute percentage error.



Step Counts

A total of 31 studies on 72 devices from 29 brands reported data on step counts. The reference standards used were manual count (directly observed or on video, usually with the help of a tally counter) or automated count through video analysis, an activity tracker (8 different devices), or a photoelectric cell.

The *ActiGraph wGT3xBT-BT*, tested against manual count, showed a mean percentage error of –41.7% (SD 13.5%) [21]. The *ActiGraph GT3x+* showed no statistically significant correlation with the same reference standard [46].

The *Apple Watch* (series not specified) was evaluated in 6% (2/31) of studies using manual count as the reference standard [30,64]. The mean difference between the device and the manual count varied from –47 (SD 470) steps to 39.44 (SD 151.81) steps in different walking conditions.

For the *Fitbit Alta*, the mean step count was 773 (SD 829) higher (P=.009) than the one obtained with the reference standard, an accelerometer [15]. For the *Fitbit Charge*, the mean difference was –59 (SD 704) steps compared with direct observation [64]. The MAPE for the same device ranged from –4.4% to 20.7%, using different automated step count methods as the reference standard [33,63,65]. The *Fitbit Charge HR* was assessed in 29% (9/31) of studies, using direct observation [19,21,28,30,61] or an automated method of step count as the reference standard [25,40,47,63]. The MAPE ranged from –12.7% to 24.1%. The accuracy of the *Fitbit Flex* in measuring steps was assessed in 35% (11/31) of studies, using manual count [13,19,21,35,36,38,39,46] or an ActiGraph device [11,51,57] as the reference standard. The mean percentage error ranged from –23% to 13%. For the *Fitbit One* and *Fitbit Zip*, no statistically significant correlation was found in step counting using direct observation as the reference standard [46]. The correlation coefficient was not reported. For the *Fitbit Surge*, the mean difference compared with direct observation was –86.0 steps (P=.004) [28]. For the *Fitbit Ultra*, the MAPE was 99.6% (SD 0.8%) [41] and the Pearson correlation coefficient against manual count ranged from 0.44 to 0.99 in different exercise conditions [58].

The accuracy of the *Garmin Vivofit* was assessed in 16% (5/31) of studies [12,13,35,53,63], with a MAPE ranging from –41% to 18% [13,53,63]. For the *Vivofit 2*, a study reported a MAPE of 4% [65] and another study reported a mean difference ranging from 5.09 (SD 8.38) steps to 98.06 (SD 137.49) steps in different walking conditions (over a maximum distance of 1.6 km) [30].

In a study by Wahl et al [63], the MAPE against automated step counting using a photoelectric cell as the reference standard, in different exercise types and conditions, ranged from –2.7% to 1.5% for the *Garmin Forerunner 920XT*, from –1.5% to 0.6% for the *Garmin Vivoactiv*, and from –1.1% to –0.3% for the *Garmin Vivosmart* [63]. For the *Garmin Vivoactive HR*, the

mean difference against manual step count was –19.7 steps (*P*=.03) [28]. For the *Garmin Vivosmart HR*, the mean difference ranged from –39.7 (SD 54.9) steps to 5.4 (SD 5.8) steps for different walking speeds and locations (outdoor vs indoor) over a total of 111-686 steps [40].

For the *Jawbone UP*, the MAPE was –6.73% in a study [65] and the mean absolute difference 806 over an average of 9959 steps in another study [29]. For the *Jawbone UP2*, the mean difference ranged from 16.19 (SD 29.14) steps to 64 (SD 66.32) steps for different walking conditions over a maximum distance of 1.6 km [30]. For the *Jawbone UP24*, the mean percentage error ranged from –28% to –0.8% [21,35,36].

For the *Misfit Shine*, the MAPE ranged from –13% to 23% [13,65].

For the *Mio Fuse*, the MAPE ranged from –5% to –16% at different treadmill speeds [25], whereas in another study, the mean percentage error was <5% for the *Xiaomi Mi Band 2* [61].

For the *Nike FuelBand*, the mean percentage error ranged from –34.3% (SD 26.8%) to –16.7% (SD 16.5%) [35], whereas for the *FuelBand SE*, the MAPE ranged from 10.2% to 45.0% [13].

The MAPE for the *Polar Loop* ranged from –13% to 27% in 3 studies [13,63,65]. Regarding 2 other devices from *Polar*, for the *A300*, a Pearson correlation coefficient of 0.96 (*P*<.01) [16] was reported, whereas for the *V800*, the Bland–Altman bias was equal to 2487 (SD 2293) steps per day over a mean 10,832 (SD 4578) steps per day measured with the reference standard [34].

For the *Withings Pulse*, the MAPE for step count ranged from –16.0% to –0.4% [63] and the accuracy from 97.2% to 99.9% [39]. All the remaining devices were only used in 1 study each, and the results are reported in Multimedia Appendix 1.

Heart Rate

A total of 9 studies on 15 devices from 7 brands evaluated the accuracy of activity-tracking devices to measure the participants' heart rates. The reference standards used were electrocardiography, pulse oximetry, or another activity tracker (4 different devices).

For the Apple Watch, the MAPE for measuring heart rate ranged from 1% (SD ~1%) to 7 (SD ~11%) [26,31].

In the *Fitbit* family of devices, for the *Fitbit Charge*, the mean bias estimated with the Bland–Altman method ranged from –6 (SD 10) bpm to –9 (SD 8) bpm [37,48,64]. For the *Fitbit Charge HR* or *Fitbit ChargeHR2*, the MAPE for measuring heart rate ranged from 2.4% (SD ~1.5%) to 17% (SD ~20%) [26,47,62]. For the *Fitbit Blaze*, the MAPE ranged from 6% (SD 6%) to 16% (SD 18%) for different activities [31].

Active Time: Time Spent in Moderate- to Vigorous-Intensity Physical Activity and Other Outcomes

A total of 13 studies on 11 devices from 8 brands reported on the time spent being active, most frequently defined as the time spent in moderate- to vigorous-intensity physical activity (MVPA; 11 studies), expressed in minutes per day. The reference standard for MVPA was another activity tracker (3 different devices). Other outcomes were time spent being active (standing+walking+running), time spent running, or time spent on different types of physical activity, with each of these outcomes being reported in only 1 study.

For the *Fitbit Flex*, the MAPE for measuring the time spent in MVPA varied from 7% (SD 6%) to 74% (SD 13%) [50] and the mean percentage error ranged from –65% to 10% [11,36]. All the other devices were only used in 1 study each, and the results are reported in Multimedia Appendix 1.

Intensity of Activity: EE and Other Outcomes

A total of 24 studies on 42 devices from 23 brands focused on measuring the intensity of physical activity. The most frequent measure of intensity was EE, expressed as kcal, evaluated in 92% (22/24) of studies. The less frequent measures of intensity included loading rate and the classification of physical activity (sedentary, household, walking, and running). For EE, the reference standard used most commonly was indirect calorimetry (6 different instruments). Less common reference standards included EE estimated with other wearable activity trackers (5 different devices), estimated based on the treadmill settings, or direct room calorimetry.

Among the *ActiGraph* family, the mean percentage difference in the EE compared with the reference standard in people with previous stroke was 3% for walking participants and 47% for participants with wheelchair using the *ActiGraph GT3X+* [24]. The Spearman correlation coefficient was 0.08 (*P*=.71) if worn on the plegic side and 0.20 (*P*=.34) if worn on the nonplegic side with the *ActiGraph GTX* [45]. Using the *ActiGraph GT1M*, the mean percentage difference was 0.5% (SD 8.0%) in a study [20], whereas another study found that the device overestimated EE at moderate intensity by 60% and underestimated EE by 40% at vigorous intensity while being 86% accurate in measuring EE at light intensity [27].

For the *Apple Watch*, the MAPE for EE ranged from 15% (SD 10%) to 211% (SD ~96%) [22,26].

In the *Fitbit* family, the MAPE from the *Charge* model ranged from –4.5% to 75.0% in different studies [22,33,63] and from –12% to 89% for the *Charge HR* [25,26,47,63]. For the *Fitbit Flex*, a mean percentage bias of –13% was reported [36]. For the *Fitbit One*, a study reported a mean bias of 2.91 (SD 4.35) kcal per minute [49], whereas for the *Fitbit Ultra*, the Pearson correlation coefficient ranged from 0.24 to 0.67 for different physical activities [58].

Among the devices from *Garmin*, the MAPE for EE ranged from –21% to 45% for the *Vivofit* [63,66], from –2% to –36% for the *Vivosmart* [63], and from 5% to 37% for the *Vivoactive* [63].

For the *Garmin Forerunner*, the MAPE ranged from –27% to 49% for the model *920XT* [52,63] and from 31% (SD ~26%) to 155% (SD ~164%) for the model 225 [26].

In the *Polar* family, the MAPE for EE ranged from 10% to 40% for the *V800* model [52], with a Bland–Altman bias of 957.5 (SD 679.9) kcal, when the mean EE measured with the reference standard was 1456.48 (SD 731.40) kcal [34]. For the *Polar Loop*, the MAPE for EE ranged from 6% to 56% [63]. The

Pearson correlation coefficient was 0.74 (P<.01) for the *Polar A300* [16].

For the *Withings Pulse*, the MAPE for EE ranged from –39% to 64% [63,66].

Outcomes Reported Less Frequently

Other outcomes that were evaluated less frequently include distance, reported in 3 studies on 15 devices from 7 brands, always using the measured distance as the reference standard [30,35,63]; speed, reported in a study using 1 device, with actual speed (on a treadmill) as the reference standard [23]; and activity count, defined as the number of activities (eg, number of arm movements or body movements based on observation or measured acceleration data), reported from 4 studies on 4 devices from 4 different brands using as the reference standard manual count (video recording), video analysis (automated), or an activity tracker [17,32,42,54].

Risk of Bias

The risk-of-bias assessment for each outcome is reported in Multimedia Appendix 1. In summary, all the studies were at high or probably high risk of bias for the domain *Patient selection* because they used a convenience sampling technique. Almost all the studies were at low risk of bias for the domains *Index test* and *Reference standard* because the 2 measurement methods were applied at the same time and interpreted without knowledge of the results obtained with the other method. A small number of studies was identified as high risk for the domain *Flow and timing* based on the high percentage (>25%) of missing data for the index test or reference standard.

Acceptability

The acceptability of wrist-wearable activity trackers was assessed in 11 studies on 10 devices from 9 brands.

Data Availability

In all, 36% (4/11) of studies focused on data availability, expressed as a proportion of time in which the data were available, and a different device was used in each of these studies. The denominator for the proportion could be the study duration or the time spent exercising. Rowlands et al [75] found that data availability was 52% in a pediatric healthy population using the GENE*Activ* for 14 months. Deka et al [68] focused on data availability during exercise time. In this study, adult patients with cardiac heart failure activated their *Fitbit Charge HR* in 75% of the exercise sessions (over 5 days) and data were available for 99% of the time when activated. Marcoux et al [72] studied the *Fitbit Flex 2* in adults with idiopathic pulmonary fibrosis (for 46 days). Of the 20 patients, 2 did not succeed in activating the device. Among the remaining participants, data were available for a mean of 91% (SD 20%) of the time. Lahti et al [71] studied the *Garmin Vivofit* in adults with schizophrenia and found data available for 97% of the time (over 4 months).

Wearing Time

In all, 27% (3/11) of studies reported on the wearing time. Farina et al [69], using the *GENEActiv*, found that 89% of the participants with dementia and 86% of their caregivers wore the device for the duration of the study (28 days). Speier et al [74], using the *Fitbit Charge 2*, enrolled participants with coronary artery disease. The median time spent wearing the activity tracker ranged from 44% to 90% over 90 days. Finally, for *Nike FuelBand***,** in a study on patients with schizophrenia, the mean wearing time was 89% (SD 13%) over 80-133 days [73].

Ease of Use and Other Characteristics

In all, 36% (4/11) of studies focused on the ease of use and similar characteristics of wrist-wearing devices. The *Polar A300* was assessed in patients with chronic obstructive pulmonary disease wearing the device for 3 days using the Post-Study System Usability Questionnaire, which calculates a score that ranges from 1 to 7 (the lower the better) for 3 subdomains [16]. The mean scores were 1.46 (SD 0.23) for system quality, 2.41 (SD 0.53) for information quality, and 3.35 (SD 0.62) for interface quality. The *AX3 data logger* was assessed in persons with Parkinson disease wearing the device for 7 days [70]. A questionnaire created ad hoc was used for the assessment; 94% of the participants agreed that they were willing to wear the sensors at home, and 85% agreed that they were willing to wear the sensors in public. However, some of the participants reported problems with the strap fitting and the material (number not reported). The *Fitbit Flex* was assessed with a questionnaire created ad hoc in a study on pregnant women followed for 7 days [57]. The *Fitbit Flex* was reported by 31% to be inconvenient, 6% to be poorly esthetic, and 12% to be uncomfortable. Kaewkannate et al [39] asked healthy participants to wear 4 different devices over 28 days and compared them using a questionnaire created ad hoc. The *Withings Pulse* had the highest satisfaction score, followed by *Misfit Shine*, *Jawbone UP24*, and *Fitbit Flex*.

Discussion

Study Findings

We systematically reviewed the available evidence on the acceptability and accuracy of wrist-wearable activity-tracking devices for measuring physical activity across different devices and measures. We found substantial heterogeneity among the included studies. The main sources of heterogeneity were the studies' population and setting, the device used, the reference standard, the outcome assessed, and the outcome measure reported.

Acceptability was evaluated in 11 studies on 10 devices from 9 brands. Data availability was ≥75% for the Fitbit Charge HR, Fitbit Flex 2, and Garmin Vivofit. Data availability is defined as the amount of data captured over a certain time period, which, in this case, is over a predetermined duration of each respective study. Data availability can be a measure of how accurate a device is at capturing data when the device is worn. For example, if an individual wears the device for 8 hours but only 4 hours of data are available, some questions may be raised on the capability of the device to capture information accurately. The wearing time was 89% for both the GENEActiv and Nike FuelBand. Wearing time is defined as the amount of time the device is worn over a predetermined duration for each study. For each study, wearing time may have been assessed differently; for example, a study may measure wearing time

over a day, whereas another study may measure over a week. Both data availability and wearing time can provide a deeper look into acceptability because participants may wear a device more frequently and, ultimately, have more data available if a device is more acceptable. Accuracy was assessed in 57 studies on 72 devices from 29 brands. Among 14 outcomes assessed, step counts, heart rate, and EE were the ones used most frequently. For step counts, the Fitbit Charge (or the Fitbit Charge HR) had a MAPE <25% across 20 studies. For heart rate, the Apple Watch had a MAPE <10% in 2 studies. For EE, the MAPE was >30% for all the brands, showing poor accuracy across devices.

Comparison With Other Systematic Reviews

Feehan et al [77] conducted a systematic review on the accuracy of Fitbit devices for measuring physical activity. The review did not specifically focus on wrist-wearable activity trackers; it also included studies using activity trackers worn on other body locations (torso, ankle, or hip). This systematic review reported a good accuracy of Fitbit devices in measuring steps, with 46% of the included studies reporting a measurement error within –3% to +3%. Regarding EE, the authors concluded that "Fitbit devices are unlikely to provide accurate measures of energy expenditure." Studies on heart rate were not included in the review. Evenson et al [78] performed a systematic review focusing on Fitbit and Jawbone devices. Similarly, wearing the device on the wrist was not an inclusion criterion. The authors concluded that for step counts, the included studies often showed a high correlation, with the correlation coefficient ≥0.80 among devices from both brands, with the reference standards. The correlation was frequently low for the outcome EE. Similar to the review by Feehan et al [77], the outcome heart rate was not included in this systematic review. The results of these systematic reviews are consistent with our findings for the devices and outcomes assessed.

Strengths and Limitations

The main strengths of our systematic review include the inclusion of all the devices reported in the literature; the reporting on all the outcomes related to acceptability and accuracy, with no restrictions; and the assessment of the risk of bias of the included studies. These characteristics make this review unique for this topic. However, in our systematic review, we decided to exclude studies in which a wearable device was not positioned on a wrist. Some devices can be positioned both on the wrist and other sites (torso, hip, ankle, arm, or brassiere), and the acceptability and accuracy can vary for the same device depending on where it is positioned, increasing heterogeneity [77,78]. Therefore, our results cannot be generalized to the acceptability and accuracy of devices worn on sites other than wrists. Acceptability is defined and measured in many different ways in the literature about wearing devices and about information technology in general [79]. These definitions are often broad and nonspecific, with published literature suggesting that acceptability research should become more robust [80]. For the purpose of our paper, acceptability was operationalized using proxies such as wearing time or data availability. However, other definitions have proposed that acceptability is related more to the extent to which individuals receiving a health care intervention find it appropriate based on cognitive and emotional responses to the intervention [80]. It is important to recognize that acceptability may be more of a holistic and subjective construct rather than an objective one, and thus wear time or data availability may not do full justice to acceptability. Although these metrics have the advantage of being relatively easy to obtain and reproduce, allowing for quantitative comparisons, they are only proxies for acceptability, which is a more nuanced concept. For example, one might wonder if wearing time is low because a person only wears the device a few hours each day or only on weekends or if they completely stopped wearing it after some time. Moreover, wearing time is more likely to offer valuable information in studies with a long follow-up, whereas 2 out of 3 studies reporting on this outcome had a follow-up of <1 week. Because of the presence of important heterogeneity among studies, we were not able to perform a meta-analysis.

Regardless, the comprehensive reporting in this review will allow researchers to assess the available evidence and inform future studies, either to further assess the accuracy of wearable devices or to inform the choice of one device over another to use in interventional studies. To facilitate these choices, we have provided to readers the database with the results of the individual included studies and we did our best to offer a synthesis of the 3 outcomes reported most frequently (step counts, heart rate, and EE).

Future Research

Further high-quality studies are needed to determine the accuracy and acceptability of wearable devices for measuring physical activity. Given the number of devices available (72 included in this review), it is unlikely that a single study will be able to answer this question. This makes it particularly important to standardize some aspects of these studies, to reduce the heterogeneity among them, and allow for meta-syntheses of the results with comparisons across studies, devices, and outcomes. If the heterogeneity was acceptable, a network meta-analysis would also allow researchers to make indirect comparisons. The main sources of heterogeneity that could be controlled are the setting of the study, the population, the reference standard used, and the outcome definition and measure. A first step in this direction would be putting together a task force of experts to issue guidelines on how to report these experiments, similar to guidelines for the EQUATOR network. A second step would be to issue recommendations on this aspect, starting with accepted reference standards against which devices should be tested for each outcome, the conditions in which the experiment should be conducted, and the way in which the outcomes should be measured and analyzed. Regarding the reference standards, some of these are more accurate than others. Our approach was to take accuracy to mean criterion and convergent validity in this review, but once there is consensus on the acceptable reference standard, other comparisons should not be included in a meta-synthesis. Regarding the method to report on the accuracy of continuous variables (more common in this field), this is the order of priority that we suggest: MAPE, mean percentage error, mean difference, Bland–Altman mean bias, and measure of correlation as the least preferred. This is because the percentage error gives the reader a better

understanding of the importance of the error (a mean error of 50 steps is much more relevant if the total step count was 100 than if it was 10,000). We preferred the MAPE over the mean absolute error because when the absolute value is not used, there is a risk of negative and positive errors balancing each other, with the risk of overestimating the accuracy. We prefer the mean difference over the Bland–Altman mean bias because in an accuracy study, the reference standard is supposed to be more accurate than the index test, and therefore the latter should be tested against the former, not against their mean. In the case of the acceptability outcome, consensus should be reached also on how to define and measure it. For example, defining a minimum set of outcomes to be reported might help in this context. This might include reporting the percentage of abandonment over time. Furthermore, as new devices become available, their acceptability and accuracy should also be tested because they could differ from the acceptability and accuracy of other devices, even those produced by the same company. Regarding the choice of the device to use in interventional studies, for example, in studies that aim at increasing physical activity in a certain population, there is no one-device-fits-all answer. This choice should be based on the available data on acceptability and accuracy and be tailored to the outcome to measure. In a study with step count as the main outcome, the Fitbit Charge and Fitbit Charge HR might be appropriate choices. The Apple Watch might be preferred if the main outcome is heart rate. Active time was most often measured through time spent in MVPA, and the Fitbit Flex is the only device that was used in 3 studies, showing good results in 2 of these. Regarding EE, we do not feel comfortable suggesting the use of any device based on the current evidence because the accuracy was poor across devices. The decision should probably be driven by the other outcomes used. Broader recommendations should be issued in the form of guidelines from a panel of experts using this systematic review as a knowledge base.

Conclusions

We reported on the acceptability and accuracy of 72 wrist-wearable devices for measuring physical activity produced by 29 companies. The Fitbit Charge and Fitbit Charge HR were consistently shown to have a good accuracy for step counts and the Apple Watch for measuring heart rate. None of the tested devices proved to be accurate in measuring EE. Efforts should be made to reduce the heterogeneity among studies.

Authors' Contributions

FG and AI conceived this review. FG is the guarantor of the review and drafted the manuscript. TN and VBD developed the search strategy. FG, NN, VBD, APB, and DP screened the articles and extracted the data. All the authors read, provided feedback, and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Database of result characteristics of the studies reporting on accuracy and acceptability.
[XLSX File (Microsoft Excel File), 34 KB - jmir_v24i1e30791_app1.xlsx]

Multimedia Appendix 2
Guide for accessing database in Multimedia Appendix 1.
[PDF File (Adobe PDF File), 206 KB - jmir_v24i1e30791_app2.pdf]

Multimedia Appendix 3
Result characteristics of the studies reporting on accuracy.
[DOCX File , 45 KB - jmir_v24i1e30791_app3.docx]

References

1. Seifert A, Schlomann A, Rietz C, Schelling HR. The use of mobile devices for physical activity tracking in older adults' everyday life. Digit Health 2017 Nov 09;3:2055207617740088 [FREE Full text] [doi: 10.1177/2055207617740088] [Medline: 29942617]
2. Ainsworth BE. How do I measure physical activity in my patients? Questionnaires and objective methods. Br J Sports Med 2009 Jan;43(1):6-9. [doi: 10.1136/bjsm.2008.052449] [Medline: 18718977]
3. Rütten A, Ziemainz H, Schena F, Stahl T, Stiggelbout M, Auweele YV, et al. Using different physical activity measurements in eight European countries. Results of the European Physical Activity Surveillance System (EUPASS) time series survey. Public Health Nutr 2003 Jun;6(4):371-376. [doi: 10.1079/PHN2002450] [Medline: 12795825]
4. Rothwell PM. Analysis of agreement between measurements of continuous variables: general principles and lessons from studies of imaging of carotid stenosis. J Neurol 2000 Nov;247(11):825-834. [doi: 10.1007/s004150070068] [Medline: 11151413]

5. Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: measures of agreement. Perspect Clin Res 2017;8(4):187-191 [FREE Full text] [doi: 10.4103/picr.PICR_123_17] [Medline: 29109937]
6. Viera AJ, Garrett JM. Understanding interobserver agreement: the kappa statistic. Fam Med 2005 May;37(5):360-363 [FREE Full text] [Medline: 15883903]
7. Dillon A, Morris M. User acceptance of information technology: theories and models. Annual Review of Information Science and Technology (ARIST). 1996. URL: https://eric.ed.gov/?id=EJ536186 [accessed 2021-12-30]
8. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: a systematic literature review. J Biomed Inform 2019 May;93:103153. [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
9. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev 2016 Dec 05;5(1):210 [FREE Full text] [doi: 10.1186/s13643-016-0384-4] [Medline: 27919275]
10. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, QUADAS-2 Group. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med 2011 Oct 18;155(8):529-536. [doi: 10.7326/0003-4819-155-8-201110180-00009] [Medline: 22007046]
11. Alharbi M, Bauman A, Neubeck L, Gallagher R. Validation of Fitbit-Flex as a measure of free-living physical activity in a community-based phase III cardiac rehabilitation population. Eur J Prev Cardiol 2016 Feb 23:1476-1485. [doi: 10.1177/2047487316634883] [Medline: 26907794]
12. Alsubheen SA, George AM, Baker A, Rohr LE, Basset FA. Accuracy of the vivofit activity tracker. J Med Eng Technol 2016 Aug;40(6):298-306. [doi: 10.1080/03091902.2016.1193238] [Medline: 27266422]
13. An H, Jones GC, Kang S, Welk GJ, Lee J. How valid are wearable physical activity trackers for measuring steps? Eur J Sport Sci 2017 Apr;17(3):360-368. [doi: 10.1080/17461391.2016.1255261] [Medline: 27912681]
14. An H, Kim Y, Lee J. Accuracy of inclinometer functions of the activPAL and ActiGraph GT3X+: a focus on physical activity. Gait Posture 2017 Jan;51:174-180 [FREE Full text] [doi: 10.1016/j.gaitpost.2016.10.014] [Medline: 27780084]
15. Blondeel A, Loeckx M, Rodrigues F, Janssens W, Troosters T, Demeyer H. Wearables to coach physical activity in patients with COPD: validity and patient's experience. Am J Respirat Crit Care Med 2018;197:A7072 [FREE Full text]
16. Boeselt T, Spielmanns M, Nell C, Storre JH, Windisch W, Magerhans L, et al. Validity and usability of physical activity monitoring in patients with chronic obstructive pulmonary disease (COPD). PLoS One 2016;11(6):e0157229 [FREE Full text] [doi: 10.1371/journal.pone.0157229] [Medline: 27305105]
17. Bruder AM, McClelland JA, Shields N, Dodd KJ, Hau R, van de Water AT, et al. Validity and reliability of an activity monitor to quantify arm movements and activity in adults following distal radius fracture. Disabil Rehabil 2018 Jun;40(11):1318-1325. [doi: 10.1080/09638288.2017.1288764] [Medline: 28637143]
18. Bulathsinghala C, Tejada J, Wallack RZ. Validating output from an activity monitor worn on the wrist in patients with COPD. Chest 2014 Oct;146(4):25A. [doi: 10.1378/chest.1989979]
19. Burton E, Hill KD, Lautenschlager NT, Thøgersen-Ntoumani C, Lewin G, Boyle E, et al. Reliability and validity of two fitness tracker devices in the laboratory and home environment for older community-dwelling people. BMC Geriatr 2018 May 03;18(1):103 [FREE Full text] [doi: 10.1186/s12877-018-0793-4] [Medline: 29724191]
20. Choi L, Chen KY, Acra SA, Buchowski MS. Distributed lag and spline modeling for predicting energy expenditure from accelerometry in youth. J Appl Physiol (1985) 2010 Mar;108(2):314-327 [FREE Full text] [doi: 10.1152/japplphysiol.00374.2009] [Medline: 19959770]
21. Chow JJ, Thom JM, Wewege MA, Ward RE, Parmenter BJ. Accuracy of step count measured by physical activity monitors: the effect of gait speed and anatomical placement site. Gait Posture 2017 Sep;57:199-203. [doi: 10.1016/j.gaitpost.2017.06.012] [Medline: 28666177]
22. Chowdhury EA, Western MJ, Nightingale TE, Peacock OJ, Thompson D. Assessment of laboratory and daily energy expenditure estimates from consumer multi-sensor physical activity monitors. PLoS One 2017;12(2):e0171720 [FREE Full text] [doi: 10.1371/journal.pone.0171720] [Medline: 28234979]
23. Cohen MD, Cutaia M. A novel approach to measuring activity in chronic obstructive pulmonary disease: using 2 activity monitors to classify daily activity. J Cardiopulm Rehabil Prev 2010;30(3):186-194. [doi: 10.1097/HCR.0b013e3181d0c191] [Medline: 20216326]
24. Compagnat M, Mandigout S, Chaparro D, Daviet JC, Salle JY. Validity of the Actigraph GT3x and influence of the sensor positioning for the assessment of active energy expenditure during four activities of daily living in stroke subjects. Clin Rehabil 2018 Dec;32(12):1696-1704. [doi: 10.1177/0269215518788116] [Medline: 30012036]
25. Dondzila C, Lewis C, Lopez J, Parker T. Congruent accuracy of wrist-worn activity trackers during controlled and free-living conditions. Int J Exerc Sci 2018;11(7):584 [FREE Full text]
26. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating accuracy at exercise intensities: a comparative study of self-monitoring heart rate and physical activity wearable devices. JMIR Mhealth Uhealth 2017 Mar 16;5(3):e34 [FREE Full text] [doi: 10.2196/mhealth.7043] [Medline: 28302596]
27. Durkalec-Michalski K, Wozniewicz M, Bajerska J, Jeszka J. Comparison of accuracy of various non-calorimetric methods measuring energy expenditure at different intensities. Hum Mov 2013;14(2):167. [doi: 10.2478/humo-2013-0019]

28. Falgoust B, Handwerger B, Rouly L, Blaha O, Pontiff M, Eason J. Accuracy of wearable activity monitors in the measurement of step count and distance. Cardiopulm Phys Ther J 2018;29(1):36 [FREE Full text]
29. Ferguson T, Rowlands AV, Olds T, Maher C. The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study. Int J Behav Nutr Phys Act 2015;12:42 [FREE Full text] [doi: 10.1186/s12966-015-0201-9] [Medline: 25890168]
30. Gaz DV, Rieck TM, Peterson NW, Ferguson JA, Schroeder DR, Dunfee HA, et al. Determining the validity and accuracy of multiple activity-tracking devices in controlled and free-walking conditions. Am J Health Promot 2018 Nov;32(8):1671-1678. [doi: 10.1177/0890117118763273] [Medline: 29558811]
31. Gillinov AM, Etiwy M, Gillinov S, Wang R, Blackburn G, Phelan D, et al. Variable accuracy of commercially available wearable heart rate monitors. J Am Coll Cardiol 2017 Mar;69(11):336. [doi: 10.1016/s0735-1097(17)33725-7]
32. Gironda RJ, Lloyd J, Clark ME, Walker RL. Preliminary evaluation of reliability and criterion validity of Actiwatch-Score. J Rehabil Res Dev 2007;44(2):223-230 [FREE Full text] [doi: 10.1682/jrrd.2006.06.0058] [Medline: 17551874]
33. Hargens TA, Deyarmin KN, Snyder KM, Mihalik AG, Sharpe LE. Comparison of wrist-worn and hip-worn activity monitors under free living conditions. J Med Eng Technol 2017 Apr;41(3):200-207. [doi: 10.1080/03091902.2016.1271046] [Medline: 28078908]
34. Hernández-Vicente A, Santos-Lozano A, De Cocker K, Garatachea N. Validation study of Polar V800 accelerometer. Ann Transl Med 2016 Aug;4(15):278 [FREE Full text] [doi: 10.21037/atm.2016.07.16] [Medline: 27570772]
35. Huang Y, Xu J, Yu B, Shull PB. Validity of FitBit, Jawbone UP, Nike+ and other wearable devices for level and stair walking. Gait Posture 2016 Dec;48:36-41. [doi: 10.1016/j.gaitpost.2016.04.025] [Medline: 27477705]
36. Imboden MT, Nelson MB, Kaminsky LA, Montoye AH. Comparison of four Fitbit and Jawbone activity monitors with a research-grade ActiGraph accelerometer for estimating physical activity and energy expenditure. Br J Sports Med 2017 May 08:844-850. [doi: 10.1136/bjsports-2016-096990] [Medline: 28483930]
37. Jo E, Lewis K, Directo D, Kim MJ, Dolezal BA. Validation of biofeedback wearables for photoplethysmographic heart rate tracking. J Sports Sci Med 2016 Sep;15(3):540-547 [FREE Full text] [Medline: 27803634]
38. Jones D, Crossley K, Dascombe B, Hart HF, Kemp J. Validity and reliability of the Fitbit Flex™ and ActiGraph GT3X+ at jogging and running speeds. Int J Sports Phys Ther 2018 Aug;13(5):860-870 [FREE Full text] [Medline: 30276018]
39. Kaewkannate K, Kim S. A comparison of wearable fitness devices. BMC Public Health 2016 May 24;16:433 [FREE Full text] [doi: 10.1186/s12889-016-3059-0] [Medline: 27220855]
40. Lamont RM, Daniel HL, Payne CL, Brauer SG. Accuracy of wearable physical activity trackers in people with Parkinson's disease. Gait Posture 2018 Jun;63:104-108. [doi: 10.1016/j.gaitpost.2018.04.034] [Medline: 29729611]
41. Lauritzen J, Muñoz A, Sevillano JL, Civit BA. The usefulness of activity trackers in elderly with reduced mobility: a case study. Stud Health Technol Inform 2013;192:759-762. [Medline: 23920659]
42. Lawinger E, Uhl T, Abel M, Kamineni S. Assessment of accelerometers for measuring upper-extremity physical activity. J Sport Rehabil 2015 Aug;24(3):236-243. [doi: 10.1123/jsr.2013-0140] [Medline: 25803521]
43. Lemmens PM, Sartor F, Cox LG, den Boer SV, Westerink JH. Evaluation of an activity monitor for use in pregnancy to help reduce excessive gestational weight gain. BMC Pregnancy Childbirth 2018 Jul 31;18(1):312 [FREE Full text] [doi: 10.1186/s12884-018-1941-8] [Medline: 30064390]
44. Magistro D, Brustio PR, Ivaldi M, Esliger DW, Zecca M, Rainoldi A, et al. Validation of the ADAMO Care Watch for step counting in older adults. PLoS One 2018;13(2):e0190753 [FREE Full text] [doi: 10.1371/journal.pone.0190753] [Medline: 29425196]
45. Mandigout S, Lacroix J, Ferry B, Vuillerme N, Compagnat M, Daviet J. Can energy expenditure be accurately assessed using accelerometry-based wearable motion detectors for physical activity monitoring in post-stroke patients in the subacute phase? Eur J Prev Cardiol 2017 Dec;24(18):2009-2016. [doi: 10.1177/2047487317738593] [Medline: 29067851]
46. Manning M, Tune B, Völgyi E, Smith W. Agreement of activity monitors during treadmill walking in teenagers with severe obesity. Pediatr Exerc Sci 2016;28:44 [FREE Full text]
47. Montoye AH, Mitrzyk JR, Molesky MJ. Comparative accuracy of a wrist-worn activity tracker and a smart shirt for physical activity assessment. Meas Phys Educ Exerc Sci 2017 Jun 08;21(4):201-211. [doi: 10.1080/1091367x.2017.1331166]
48. Powierza CS, Clark MD, Hughes JM, Carneiro KA, Mihalik JP. Validation of a self-monitoring tool for use in exercise therapy. PM R 2017 Nov;9(11):1077-1084 [FREE Full text] [doi: 10.1016/j.pmrj.2017.03.012] [Medline: 28400221]
49. Price K, Bird SR, Lythgo N, Raj IS, Wong JY, Lynch C. Validation of the Fitbit One, Garmin Vivofit and Jawbone UP activity tracker in estimation of energy expenditure during treadmill walking and running. J Med Eng Technol 2017 Apr;41(3):208-215. [doi: 10.1080/03091902.2016.1253795] [Medline: 27919170]
50. Redenius N, Kim Y, Byun W. Concurrent validity of the Fitbit for assessing sedentary behavior and moderate-to-vigorous physical activity. BMC Med Res Methodol 2019 Feb 07;19(1):29 [FREE Full text] [doi: 10.1186/s12874-019-0668-1] [Medline: 30732582]
51. Reid RE, Insogna JA, Carver TE, Comptour AM, Bewski NA, Sciortino C, et al. Validity and reliability of Fitbit activity monitors compared to ActiGraph GT3X+ with female adults in a free-living environment. J Sci Med Sport 2017 Jun;20(6):578-582. [doi: 10.1016/j.jsams.2016.10.015] [Medline: 27887786]

52. Roos L, Taube W, Beeler N, Wyss T. Validity of sports watches when estimating energy expenditure during running. BMC Sports Sci Med Rehabil 2017 Dec 20;9(1):22 [FREE Full text] [doi: 10.1186/s13102-017-0089-6] [Medline: 29296281]
53. Schaffer SD, Holzapfel SD, Fulk G, Bosch PR. Step count accuracy and reliability of two activity tracking devices in people after stroke. Physiother Theory Pract 2017 Oct;33(10):788-796. [doi: 10.1080/09593985.2017.1354412] [Medline: 28777710]
54. Scott JJ, Rowlands AV, Cliff DP, Morgan PJ, Plotnikoff RC, Lubans DR. Comparability and feasibility of wrist- and hip-worn accelerometers in free-living adolescents. J Sci Med Sport 2017 Dec;20(12):1101-1106. [doi: 10.1016/j.jsams.2017.04.017] [Medline: 28501418]
55. Semanik P, Lee J, Pellegrini CA, Song J, Dunlop DD, Chang RW. Comparison of physical activity measures derived from the Fitbit Flex and the ActiGraph GT3X+ in an employee population with chronic knee symptoms. ACR Open Rheumatol 2020 Jan 02;2(1):48-52 [FREE Full text] [doi: 10.1002/acr2.11099] [Medline: 31943970]
56. Sirard JR, Masteller B, Freedson PS, Mendoza A, Hickey A. Youth oriented activity trackers: comprehensive laboratory- and field-based validation. J Med Internet Res 2017 Jul 19;19(7):e250 [FREE Full text] [doi: 10.2196/jmir.6360] [Medline: 28724509]
57. St-Laurent A, Mony MM, Mathieu M, Ruchat SM. Validation of the Fitbit Zip and Fitbit Flex with pregnant women in free-living conditions. J Med Eng Technol 2018 May;42(4):259-264. [doi: 10.1080/03091902.2018.1472822] [Medline: 30198806]
58. Stackpool CM. Accuracy of various activity trackers in estimating steps taken and energy expenditure. UW-L Master's Theses. 2013. URL: https://minds.wisconsin.edu/handle/1793/66292 [accessed 2021-12-31]
59. Stiles VH, Griew PJ, Rowlands AV. Use of accelerometry to classify activity beneficial to bone in premenopausal women. Med Sci Sports Exerc 2013 Dec;45(12):2353-2361. [doi: 10.1249/MSS.0b013e31829ba765] [Medline: 23698245]
60. Støve MP, Haucke E, Nymann ML, Sigurdsson T, Larsen BT. Accuracy of the wearable activity tracker Garmin Forerunner 235 for the assessment of heart rate during rest and activity. J Sports Sci 2019 Apr;37(8):895-901. [doi: 10.1080/02640414.2018.1535563] [Medline: 30326780]
61. Tam KM, Cheung SY. Validation of electronic activity monitor devices during treadmill walking. Telemed J E Health 2018 Oct;24(10):782-789. [doi: 10.1089/tmj.2017.0263] [Medline: 29364065]
62. Thomson EA, Nuss K, Comstock A, Reinwald S, Blake S, Pimentel RE, et al. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. J Sports Sci 2019 Jun;37(12):1411-1419. [doi: 10.1080/02640414.2018.1560644] [Medline: 30657025]
63. Wahl Y, Düking P, Droszez A, Wahl P, Mester J. Criterion-validity of commercially available physical activity tracker to estimate step count, covered distance and energy expenditure during sports conditions. Front Physiol 2017 Sep;8:725 [FREE Full text] [doi: 10.3389/fphys.2017.00725] [Medline: 29018355]
64. Wallen MP, Gomersall SR, Keating SE, Wisløff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11(5):e0154420 [FREE Full text] [doi: 10.1371/journal.pone.0154420] [Medline: 27232714]
65. Wang L, Liu T, Wang Y, Li Q, Yi J, Inoue Y. Evaluation on step counting performance of wristband activity monitors in daily living environment. IEEE Access 2017;5:13020-13027. [doi: 10.1109/access.2017.2721098]
66. Woodman JA, Crouter SE, Bassett DR, Fitzhugh EC, Boyer WR. Accuracy of consumer monitors for estimating energy expenditure and activity type. Med Sci Sports Exerc 2017 Feb;49(2):371-377. [doi: 10.1249/MSS.0000000000001090] [Medline: 27580155]
67. Zhang S, Murray P, Zillmer R, Eston RG, Catt M, Rowlands AV. Activity classification using the GENEA: optimum sampling frequency and number of axes. Med Sci Sports Exerc 2012 Nov;44(11):2228-2234. [doi: 10.1249/MSS.0b013e31825e19fd] [Medline: 22617400]
68. Deka P, Pozehl B, Norman JF, Khazanchi D. Feasibility of using the Fitbit Charge HR in validating self-reported exercise diaries in a community setting in patients with heart failure. Eur J Cardiovasc Nurs 2018 Oct;17(7):605-611. [doi: 10.1177/1474515118766037] [Medline: 29546995]
69. Farina N, Sherlock G, Thomas S, Lowry RG, Banerjee S. Acceptability and feasibility of wearing activity monitors in community-dwelling older adults with dementia. Int J Geriatr Psychiatry 2019 Apr;34(4):617-624. [doi: 10.1002/gps.5064] [Medline: 30701592]
70. Fisher JM, Hammerla NY, Rochester L, Andras P, Walker RW. Body-worn sensors in Parkinson's disease: evaluating their acceptability to patients. Telemed J E Health 2016 Jan;22(1):63-69 [FREE Full text] [doi: 10.1089/tmj.2015.0026] [Medline: 26186307]
71. Lahti A, White D, Katiyar N, Pei H, Baker S. 325. Clinical utility study towards the use of continuous wearable sensors and patient reported surveys for relapse prediction in patients at high risk of relapse in schizophrenia. Biolog Psychiatry 2017 May;81(10):S133. [doi: 10.1016/j.biopsych.2017.02.340]
72. Marcoux V, Fell C, Johannson K. Daily activity trackers and home spirometry in patients with idiopathic pulmonary fibrosis. Am J Respirat Crit Care Med 2018;197:A1586 [FREE Full text]
73. Naslund JA, Aschbrenner KA, Barre LK, Bartels SJ. Feasibility of popular m-health technologies for activity tracking among individuals with serious mental illness. Telemed J E Health 2015 Mar;21(3):213-216. [doi: 10.1089/tmj.2014.0105] [Medline: 25536190]

74. Speier W, Dzubur E, Zide M, Shufelt C, Joung S, Van Eyk JE, et al. Evaluating utility and compliance in a patient-based eHealth study using continuous-time heart rate and activity trackers. J Am Med Inform Assoc 2018 Oct 01;25(10):1386-1391. [doi: 10.1093/jamia/ocy067] [Medline: 29850807]
75. Rowlands AV, Harrington DM, Bodicoat DH, Davies MJ, Sherar LB, Gorely T, et al. Compliance of adolescent girls to repeated deployments of wrist-worn accelerometers. Med Sci Sports Exerc 2018 Jul;50(7):1508-1517. [doi: 10.1249/MSS.0000000000001588] [Medline: 29474208]
76. Noun Project: Free Icons & Stock Photos for Everything. URL: https://thenounproject.com/ [accessed 2021-09-11]
77. Feehan LM, Geldman J, Sayre EC, Park C, Ezzat AM, Yoo JY, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth 2018 Aug 09;6(8):e10527 [FREE Full text] [doi: 10.2196/10527] [Medline: 30093371]
78. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
79. Nadal C, Doherty G, Sas C. Technology acceptability, acceptance and adoption-definitions and measurement. In: Proceedings of the CHI 2019 Conference on Human Factors in Computing Systems. 2019 Presented at: CHI 2019 Conference on Human Factors in Computing Systems; May 4–9, 2019; Glasgow, Scotland UK URL: https://www.researchgate.net/publication/331787774_Technology_acceptability_acceptance_and_adoption_-_definitions_and_measurement
80. Sekhon M, Cartwright M, Francis JJ. Acceptability of healthcare interventions: an overview of reviews and development of a theoretical framework. BMC Health Serv Res 2017 Jan 26;17(1):1-13 [FREE Full text] [doi: 10.1186/s12913-017-2031-8] [Medline: 28126032]

Abbreviations

EE: energy expenditure
MAPE: mean absolute percentage error
MVPA: moderate to vigorous-intensity physical activity

Edited by A Mavragani; submitted 28.05.21; peer-reviewed by F Yu, M Western; comments to author 30.07.21; revised version received 24.09.21; accepted 06.12.21; published 21.01.22.

Please cite as:
Germini F, Noronha N, Borg Debono V, Abraham Philip B, Pete D, Navarro T, Keepanasseril A, Parpia S, de Wit K, Iorio A
Accuracy and Acceptability of Wrist-Wearable Activity-Tracking Devices: Systematic Review of the Literature
J Med Internet Res 2022;24(1):e30791
URL: https://www.jmir.org/2022/1/e30791
doi:10.2196/30791
PMID:35060915

Review

Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials

Alexandre Mazeas[1,2,3], MSc; Martine Duclos[2,4], PhD; Bruno Pereira[5], PhD; Aïna Chalabaev[1], PhD

[1]Univ. Grenoble Alpes, SENS, 38000 Grenoble, France

[2]National Research Institute for Agriculture, Food and Environment (INRAE), 63000 Clermont-Ferrand, France

[3]Kiplin, 44200 Nantes, France

[4]Department of Sport Medicine and Functional Exploration, University Hospital Clermont-Ferrand, Hospital G. Montpied, 63000 Clermont-Ferrand, France

[5]Department of Biostatistics unit, Clermont-Ferrand University Hospital, 63000 Clermont-Ferrand, France

Corresponding Author:
Alexandre Mazeas, MSc
Univ. Grenoble Alpes
SENS
1741 rue de la Piscine
38000 Grenoble
France
Phone: 33 4 76 63 50 81
Email: alexandre.mazeas@univ-grenoble-alpes.fr

Abstract

Background: Gamification refers to the use of game elements in nongame contexts. The use of gamification to change behaviors and promote physical activity (PA) is a promising avenue for tackling the global physical inactivity pandemic and the current prevalence of chronic diseases. However, there is no evidence of the effectiveness of gamified interventions with the existence of mixed results in the literature.

Objective: The aim of this systematic review and meta-analysis is to evaluate the effectiveness of gamified interventions and their health care potential by testing the generalizability and sustainability of their influence on PA and sedentary behavior.

Methods: A total of 5 electronic databases (PubMed, Embase, Scopus, Web of Science, and the Cochrane Central Register of Controlled Trials) were searched for randomized controlled trials published in English from 2010 to 2020. Eligibility criteria were based on the components of the participants, interventions, comparators, and outcomes framework. Studies were included when they used gamified interventions in daily life with an active or inactive control group and when they assessed a PA or sedentary behavior outcome. We conducted meta-analyses using a random-effects model approach. Sensitivity analyses, influence analyses, and publication bias analyses were performed to examine the robustness of our results.

Results: The main meta-analysis performed on 16 studies and 2407 participants revealed a small to medium summary effect of gamified interventions on PA behavior (Hedges g=0.42, 95% CI 0.14-0.69). No statistical difference among different subgroups (adults vs adolescents and healthy participants vs adults with chronic diseases) and no interaction effects with moderators such as age, gender, or BMI were found, suggesting good generalizability of gamified interventions to different user populations. The effect was statistically significant when gamified interventions were compared with inactive control groups, such as waiting lists (Hedges g=0.58, 95% CI 0.08-1.07), and active control groups that included a nongamified PA intervention (Hedges g=0.23, 95% CI 0.05-0.41). This suggests that gamified interventions are not only efficient in changing behavior but also more effective compared with other behavioral interventions. The long-term effect (measured with follow-up averaging 14 weeks after the end of the intervention) was weaker, with a very small to small effect (Hedges g=0.15, 95% CI 0.07-0.23).

Conclusions: This meta-analysis confirms that gamified interventions are promising for promoting PA in various populations. Additional analyses revealed that this effect persists after the follow-up period, suggesting that it is not just a novelty effect caused by the playful nature of gamification, and that gamified products appear effective compared with equivalent nongamified PA interventions. Future rigorous trials are required to confirm these findings.

(J Med Internet Res 2022;24(1):e26779) doi:10.2196/26779

KEYWORDS

behavior change; eHealth; gamification; health behavior; intervention; meta-analysis; mobile phone; physical activity; systematic review

Introduction

Background

Physical inactivity and sedentary behavior (SB) are among the leading risk factors for global mortality [1]. Each year, physical inactivity is responsible for >5 million deaths worldwide [2]. In contrast, regular physical activity (PA) prevents the risk of developing chronic diseases [3,4], limits their progression [5,6], and reduces early mortality [7]. In parallel, there is a dose-response relationship between total sedentary time per day and overall mortality [7]. Meta-analyses demonstrate that the risk of mortality in adults increases steadily with a sedentary lifestyle of >3 hours per day and more significantly when this time exceeds 7 hours per day [8]. However, recent studies have suggested that high levels of PA could attenuate or even eliminate the deleterious effects of SB on overall mortality [9].

In this context, it is urgent to develop interventions that can effectively change PA. Therefore, digital health interventions constitute a new opportunity to take care of patients by involving them in their treatment in a dynamic and interactive way. Gamification is a promising avenue to capitalize on the efficacy of digital interventions. Gamification is defined as the use of game design elements in nongame contexts [10]. By integrating game mechanisms in interventions that are initially devoid of them, the purpose of gamification is to integrate into daily life the ingredients that make games enjoyable to motivate participants to engage in PA [11]. The use of *motivational affordances* created by gamification can influence psychological (eg, motivation, attitude, and enjoyment) and physical outcomes (eg, physical capacities) [12] and therefore appears as a potentially powerful technique for behavior change.

By gamifying PA, participants are encouraged to move and walk to play, which tends to make their activities more enjoyable and playful [13]. Unlike serious games, which refer to the use of a full-fledged video game for educational or health purposes (ie, a video game in its entirety as opposed to selected elements or individual features of a game) [10] and require a dedicated time, a location, and implementation [14], gamification techniques are relatively open to varying situational modes of engagement [10] and concern instead global PA in all aspects of daily life (eg, walking, running, or gardening). Gamification is made possible by mobile technologies and wearable devices that can track and collect daily activities in a continuous and web-based manner. This allows for intervening directly on the lifestyle of individuals without adding material or time constraints for the participants.

However, several literature reviews [13,15,16] have reported inconsistent results concerning the use of gamification in behavioral interventions, with some studies demonstrating positive effects and other studies providing more mixed effects. These reviews also emphasized the lack of high-quality studies and highlighted the need for more rigorous trials to isolate the impact of gamification (ie, randomized controlled trials [RCTs]). Importantly, Koivisto and Hamari [15] suggested that the effects of gamification could be smaller when using rigorous experimentation. In sum, these reviews indicate that there is no clear evidence of the effectiveness of gamified interventions. Nevertheless, no meta-analysis has been conducted yet. Quantifying the effect size of gamified interventions and identifying moderators of this effect would provide important information about the effectiveness of such interventions. Moreover, a meta-analysis appears as timely, as there are now enough RCTs to conduct such an analysis.

This Study

This study is the first to quantify the effects of gamified interventions on PA. Beyond the effect during or just after the intervention, we also seek to evaluate the long-term effects to determine the health relevance of these interventions. Indeed, we reasoned that to be considered effective, gamification must sustain its impacts over the long term and offer more than a short-term novelty effect [11]. The generalizability of gamification to different user populations is also a major issue because it would determine whether gamification can be introduced in health care settings with patients or it is more suited in prevention for healthy audiences.

The objectives of this systematic review and meta-analysis are to answer these research gaps by (1) evaluating the effect of gamified interventions on PA and SB, (2) assessing the long-term or sustained effects of gamified programs, and (3) evaluating the generalizability of gamification across different populations.

Methods

Design

This review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines [17]. Following recommendations to minimize bias and provide evidence of a priori analysis intentions [18], the study was preregistered under the international prospective register of systematic reviews (PROSPERO; registration number: CRD42020186882) and on the Open Science Framework (OSF) [19]. Moreover, all materials and data are available on the OSF page of the project to facilitate reproducibility and transparency of this review [20].

Search Strategy and Information Sources

We conducted a systematic literature search using five electronic databases: PubMed, Embase, Scopus, Web of Science, and the Cochrane Central Register of Controlled Trials. We combined alternate terms and keywords representing synonyms for the outcomes (PA or SB), intervention (gamification), device (eHealth), and trial (see Table S1 in Multimedia Appendix 1

[21-36] for an overview of the terms used). The search strategy was reviewed by an academic librarian. All databases were searched individually from January 1, 2010 (2010 being the date of the widespread adoption of the term *gamification* [10]), to December 31, 2020, and the research was restricted to English-language texts. The complete search equations for all databases are available in Multimedia Appendix 1. In addition, we complemented our search with reference harvesting from the included studies and overview articles.

Eligibility Criteria

Studies were eligible for inclusion if they were RCTs and if they met other criteria based on the participants, interventions, comparators, and outcomes framework.

Participants

This review focused on the general population regardless of age, gender, or health status (ie, patients with chronic diseases were also included). We excluded studies involving participants with contraindications to PA or with diseases preventing them from engaging in PA or understanding the principles of the game (intellectual and cognitive impairments).

Intervention

Digital interventions targeting PA or SB and incorporating game elements and gamification techniques, such as points, levels, rewards, leaderboards, narratives, and teams, were of interest. We clearly distinguished between gamification and related constructs, such as serious games. Therefore, we excluded interventions based on active video games (ie, electronic games that allow players to physically play with the images on screen) that are more comparable with serious games than with gamified products.

Comparators

Studies that attempted to compare gamified interventions with control groups without gamification elements in a randomized design were integrated in the review. These groups could be either inactive (nonexposed control group, such as a waiting list) or active (another nongamified intervention).

Outcomes

In this review, we included studies assessing change in total PA or leisure PA (quantity in metabolic equivalent of task [MET] hour per week or MET minute per week or in duration, energy expenditure [METs], moderate to vigorous PA [MVPA], step count, walking time, and active minutes) and change in time spent in SB (total time, leisure time, work, time spent in front of the computer, and time spent in front of television). These outcomes were continuous data either objectively measured (through accelerometers, pedometers, and smartphones) or subjectively measured by self-reported questionnaires. Data measured objectively were always prioritized in the analyses over self-reported questionnaires, which are more susceptible to bias with a potential overestimation of PA [37].

In addition, studies were excluded if they came from a review, commentary, or conference abstract; if they included data previously published in another study; if they were not randomized and controlled; if they were not written in English; and if they were published before 2010.

Screening

In total, 2 authors (AM and AC) independently screened the titles and abstracts resulting from the search. Full texts of the potential included studies were checked before inclusion. Disagreements were resolved by discussion or by consulting a third author (MD), and agreement was measured using the κ statistic. A complete list of excluded studies is available on the OSF page of the project.

Data Extraction

In the data collection process, AM and AC extracted data independently and were blinded to each other using a predetermined and tested template. Disagreements were resolved by discussion or consultation with a third author (MD). Extracted data included the results of each study and three types of potential moderators:

1. Population-level moderators to assess the generalizability of the intervention (population characteristics, age, gender, and pathology).
2. Intervention-level moderators to better understand gamification mechanisms (theoretical model used to develop the intervention, gamification elements, and modality of the intervention [eg, internet-based, smartphone app, and presence of social incentives]).
3. Outcome-level moderators (outcomes, measure of PA or SB, and device).

Risk of Bias Assessment

For each eligible study, 2 reviewers (AM and AC) assessed the risk of bias using the purpose-built Cochrane risk of bias tool (Table 8.5 in the Cochrane Handbook for Systematic Reviews of Interventions [38]), which evaluates 7 domains (sequence generation, allocation concealment, blinding of participants and personnel, blinding of outcome assessment, incomplete outcome data, selective outcome reporting, and other risk of bias). A judgment of the potential risk of bias was made from the extracted information and rated as *high risk*, *low risk*, or *unclear* if the related information was not available. These evaluations of bias are reported in the review and included in the analysis, and a measure of agreement with the κ statistic was calculated. After the full assessment, we decided not to present the item *blinding of participants and personnel* in the review because it was similar for all studies, which were rated as *high risk*, blinding being unfeasible for this kind of intervention.

Data Synthesis and Statistical Analyses

First, data were synthesized in a qualitative review assessing the key elements of the studies and highlighting intervention differences. This qualitative review integrates all studies that met the eligibility criteria, including those for which we were unable to extract the data.

For the quantitative analysis, means and SDs of continuous PA or SB outcomes from individual studies were compiled when available or estimated using the method by Hozo et al [39] when median and IQR were reported. When the necessary data were not available in the original article, we first requested them from

the authors. If data could not be obtained, we extracted them from the graphs when available. If this was not possible, we excluded the study from the quantitative analysis.

A global meta-analysis was conducted to obtain a summary effect. In addition, when sufficient data were available (ie, 4 studies or more reporting an outcome), we conducted different meta-analyses for each specific outcome (steps, MVPA, and time in SB) and for the follow-up effect. To address the nonindependence of data caused by study effect, random-effects models [40] were preferred to the usual statistical tests. In addition, the Hartung–Knapp–Sidik–Jonkman method was used to reduce the production of false positives inherent to the DerSimonian–Laird method [41] and to obtain more robust estimates of variance. Continuous outcomes were analyzed using standardized mean difference (SMD) to account for different measurement instruments or mean difference (MD) when the measurements were close enough. We computed Hedges *g* [42] for effect sizes, which is similar to Cohen *d* but corrects for small sample bias, which are recurrent in the studies included. Thus, a Hedges *g* of 0.2 represents a small effect; 0.5, a moderate effect; and 0.8, a large effect [43]. We computed SMDs for outcome scores after the intervention (presented in the review) and change-from-baseline outcomes. Scores on postintervention effect sizes refer to treatment group results compared with the control group results after interventions. Change-from-baseline score effect sizes were calculated as a comparison between the treatment group pre–post effect size and control group pre–post effect size.

For studies that included multiple outcomes, we kept in the main analysis the primary outcome targeted in the initial article. If none of the PA outcomes reported by a study were the primary ones, we selected the one that was the most relevant from the perspective of the intervention and the original experiment. In designs with multiple time measurements, the assessment that was the most proximal to the end of the intervention was conserved. A time assessment had to be performed >2 weeks after the end of the intervention to be included in the follow-up analysis. When studies included multiple intervention groups with gamification features, they were combined into one group following the formulae recommended by the Cochrane Handbook [38]. Studies including multiple control groups could be integrated into different subgroup analyses if they compared their gamified intervention to both an active and an inactive control group.

Statistical heterogeneity was tested using forest plots and the I^2 statistic, which is the most common metric for measuring the magnitude of between-study heterogeneity and is easily interpretable (0%-40% might not be important, 30%-60% may represent moderate heterogeneity, 50%-90% may represent substantial heterogeneity, and 75%-100% may represent considerable heterogeneity) [44]. We conducted different influential analyses to address between-study heterogeneity. We first explored the presence of outliers, defined as studies with CIs that do not overlap with the CI of the summary effect. We also performed leave-one-out analyses, which recalculated the summary effect several times, with 1 study omitted each time. Finally, we performed a Baujat plot [45], which is a diagnostic plot to detect sources of heterogeneity in the meta-analysis by comparing the contribution of each trial in the pooled effect with the overall Cochran *Q* test for heterogeneity.

We applied different methods to detect publication bias (funnel plot, Egger regression test [46], and Duval and Tweedie trim-and-fill procedure [47]). In addition, another approach to determine the evidential value of studies included in the analysis is to check the statistical power of individual studies. Therefore, we performed a sunset funnel plot [48], which is a funnel plot variant that visualizes the statistical power of each study included in the meta-analysis based on the summary effect size.

Thus, sensitivity analyses were conducted to address studies with a high risk of bias or a strong heterogeneity in the sample or studies identified as outliers. Subgroup analyses were conducted to explore possible sources of heterogeneity and test for population differences. Therefore, we conducted tests for subgroup differences using a random-effects model. In addition, moderation analyses were performed to explore the impact of potential explanatory variables and moderators on the effect size with meta-regressions when sufficient data were available (ie, at least 10 studies for each explanatory variable [38]). The results were expressed as regression coefficient estimates, 95% CIs, and *P* values.

For crossover trials, we first checked whether carry-over or period effects were problematic in the original texts of studies. For cluster randomized trials, we checked if the influence of the different clusters was not too important, analyzing the values of the intraclass correlation coefficient in the studies. Then, in the absence of sufficient information in the published articles, we addressed these studies as traditional parallel trials. Nevertheless, this procedure increased the probability of a unit of analysis error. Therefore, we performed sensitivity analyses without clusters and crossover trials to test the robustness of our results.

A summary of the analytical procedure is available in Multimedia Appendix 1 (Figure S1). Analyses were performed on R (The R Project for Statistical Computing) using the *dmetar* package [49]. Risk of bias summary and risk of bias graphs were made via the *robvis* R package [50].

Results

Characteristics of Included Studies

We screened the titles and abstracts of 1626 articles, and 51 full-text articles were assessed for eligibility according to the inclusion criteria. Finally, 18 articles [21-36,51,52] were included in the qualitative analysis and 16 were included in the meta-analysis (Figure 1). The κ value of agreement for the screening phase was 0.64 between the 2 authors, reflecting good agreement [53].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flowchart of the literature search and screening process. PA: physical activity; RCT: randomized controlled trial.

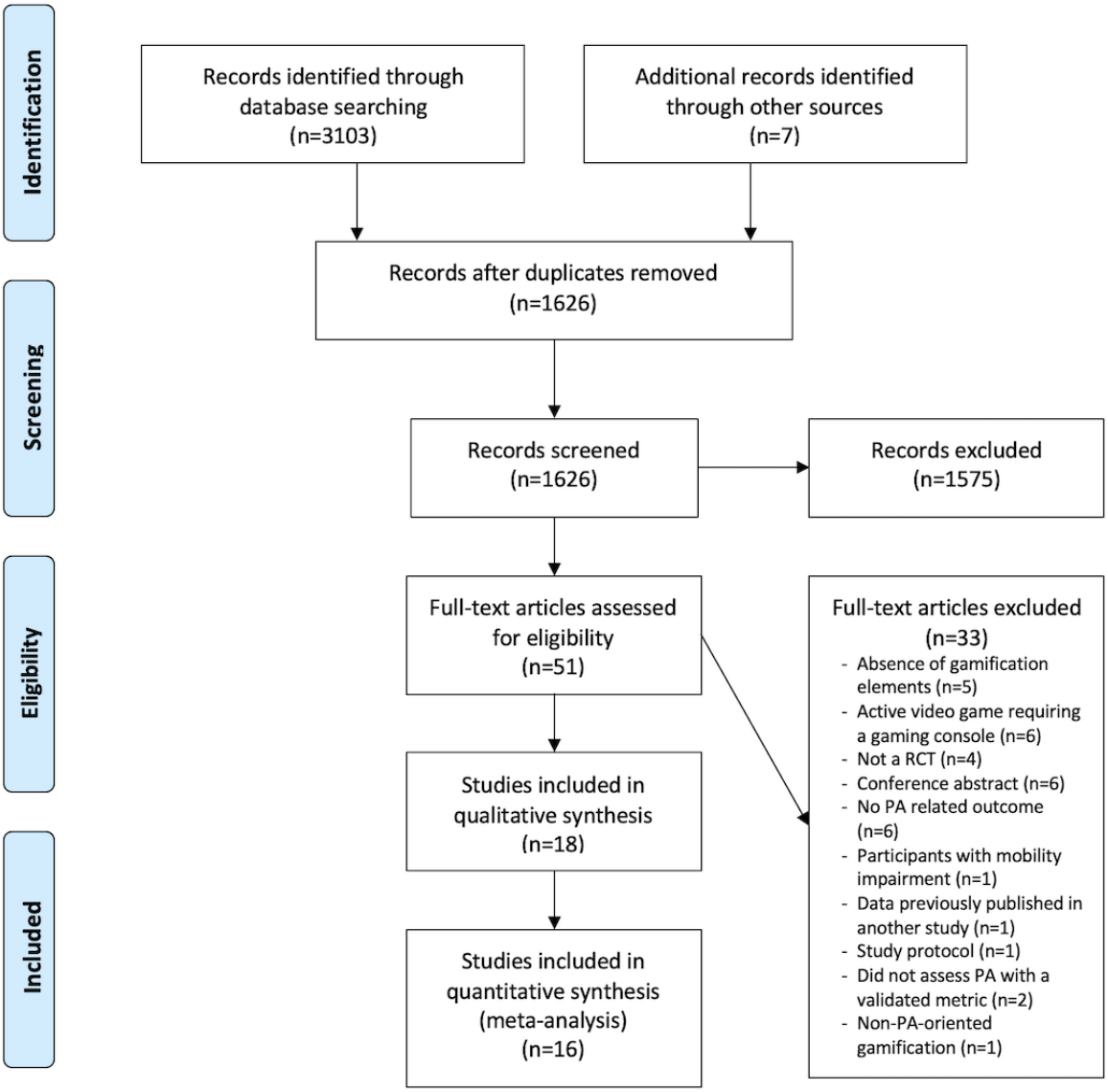


Table 1 provides an overview of the characteristics of the included studies. The 16 studies included in the quantitative analysis involved 2407 participants aged 9-73 years (mean 35.7 years, SD 17.2 years), with sample sizes ranging from 20 to 602. Overall, 67% (12/18) of studies included adult participants [22,24-30,32,34-36] and 33% (6/18) of studies included adolescents (ie, <18 years [21,23,31,33,51,52]). A total of 22% (4/18) of studies included patients with chronic diseases (ie, obesity [27,36], type 2 diabetes [35], and cardiovascular disease [28]).

A total of 6 trials were conducted in Europe; 4 in the United States; 3 in Australia and New Zealand; 3 in Canada; and 2 in Asia. Studies were published between 2014 and 2020, with 39% (7/18) published after 2018.

Most studies were based on a smartphone app (n=10; [23,28,30-35,51,52]), web-based (n=3; [21,24,29]), or both (n=4; [22,26,27,36]). Nishiwaki et al [25] used a pedometer with computerized game functions. The duration of the intervention varied from 1 to 24 weeks, with a mean of 11.8 weeks, and the most common length was 24 weeks. The most used game mechanics were internet-based rewards, such as badges, medals, or trophies (13/18, 72%; [21,22,24,26-28,30,32,33,35,36,51,52]), teams or leagues (13/18, 72%; [21-24,26,27,29,32-34,36,51,52]), levels (9/18, 50%; [22,26,27,29,33,35,36,51,52]), points or scores (7/18, 39%; [22,26,27,29,30,35,36]), or the presence of a leaderboard (7/18, 39%; [22,29,30,33,34,51,52]). Almost all studies included social incentives such as team collaboration, social networking, and messaging facilities in their intervention (15/18, 83%; all except Nishiwaki et al [25], Direito et al [31], and Höchsmann et al [35]).

Table 1. Characteristics of included studies.

Study	Participants	Intervention	Theory	PA[a] outcomes
Corepal et al [21][b]	Adolescents, n=224 (aged 12-14 years; 47% male participants)	The "StepSmart Challenge" was a web-based intervention that used gamification strategies to encourage and support PA behavior change; duration: 22 weeks	SDT[c]	Daily step count and MVPA[d] (min/day) objectively measured (Actigraph GT3x)
Dadaczynski et al [22][b]	Adult workers in an automobile manufacture, n=144 (65% male participants)	"Healingo Fit" had the objective to promote low levels of PA using a tracking-based approach measuring PA with a Fitbit pedometer and a gamified intervention accessible by desktop and mobile devices; duration: 6 weeks	SCT[e], TPB[f], and health action process approach	Self-reported VPA[g], MPA[h], and minutes walked (min/week; IPAQ[i])
Direito et al [31][b]	Adolescents, n=35 (mean age 15.7 years, SD 1.2 years; 45% male participants; BMI 22.85)	"Zombies, run! 5K Training app" was a fully automated training program designed to improve fitness, combined with an immersing and fun story; duration: 8 weeks	SDT	MVPA, VPA, MPA, and LPA[j] and sedentary time (min/day) objectively measured (Actigraph GT3x) and self-reported PA (PAQ-A[k])
Edney et al [32][b]	Adults, n=284 (mean age 41.2 years, SD 11.2 years; 25% male participants; BMI 30.1)	"Active Team" was a mobile app designed to encourage inactive adults to meet PA guidelines. Gamification and social features were implemented to increase the social comparison, support, and influence among participants; duration: 12 weeks	SCT	MVPA (min/day) objectively measured (GENEActiv) and self-reported PA (Active Australia Survey)
Garde et al [51]	Adolescents, n=47 (mean age 10.3 years, SD 1.9 years; 34% male participants; BMI *z*-score 0.35)	"MobileKids Monster Manor" was a mobile exergame synchronized with an external activity monitor. The overall goal was to complete the story with PA and steps; duration: 1 week	SDT	Daily step count and active min/day objectively measured (Tractivity)
Garde et al [33][b]	Adolescents, n=56 (mean age 11.3 years, SD 1.2 years; 62% male participants; BMI *z*-score 0.28)	"MobileKids Monster Manor" was a mobile exergame synchronized with an external activity monitor. The overall goal was to complete the story with PA and steps; duration: 1 week	SDT	Daily step count and active min/day objectively measured (Tractivity)
Garde et al [52]	Adolescents, n=37 (mean age 10.6 years, SD 0.5 years; 43% male participants; BMI *z*-score 0.21)	"MobileKids Monster Manor" was a mobile exergame synchronized with an external activity monitor. The overall goal was to complete the story with PA and steps; duration: 2 weeks	SDT	Daily step count and active min/day objectively measured (Tractivity)
Gremaud et al [34][b]	Adult office workers, n=144 (mean age 40.5 years, SD 11.4 years; 76% male participants; BMI 29.7)	"MapTrek" was a mobile health platform that gamified Fitbit use for promoting PA by placing users in a series of internet-based walking races; duration: 10 weeks	SCT	Daily step count and daily active minutes count objectively measured (Fitbit Zip activity monitor)
Höchsmann et al [35][b]	Patients with type 2 diabetes mellitus and obesity, n=35 (mean age 58.5 years; 53% male participants; BMI 32)	The intervention was a mobile app including a storyline, virtual rewards, individualized exercises, and daily PA promotion through a game; duration: 24 weeks	Taxonomy of behavior change techniques	Daily step count objectively measured (Garmin Vivofit 2)
Kurtzman et al [36][b]	Adults with obesity, n=196 (mean age 41.4 years, SD 12.2 years; 13% male participants; BMI 36.2)	Participants were in teams of 2 and had to complete weekly goal targets to win points and badges; duration: 24 weeks	Behavioral economics	Mean step count objectively measured (Withings wrist-worn device)
Leinonen et al [23][b]	Adolescents, n=496 (mean age 17.8 years, SD 0.6 years; 100% male participants; BMI 23.1)	The intervention was an app proposing a mixed-reality conquering game in which physical and social activities are rewarded; duration: 24 weeks	Transtheoretical model	Daily MVPA and daily sedentary time objectively measured (Polar Active)

Study	Participants	Intervention	Theory	PA[a] outcomes
Maher et al [24][b]	Adults, n=110 (mean age 35.6 years, SD 12.4 years; 42% male participants)	"Active Team" was a Facebook (Meta Platforms) app designed to encourage inactive adults to meet PA guidelines. Gamification and social features were implemented to increase the social comparison, support, and influence among participants; duration: 8 weeks	TPB	Self-reported MVPA, VPA, MPA, and minutes walked (min/week; Active Australia Survey)
Nishiwaki et al [25][b]	Adults, n=20 (mean age 31 years, SD 3 years; 30% male participants; BMI 21.5)	Participants wore an activity monitor with computerized game functions, such as a story, a character, and objectives; duration: 6 weeks	—[l]	Daily step count and MVPA (metabolic equivalent of tasks hour/day) objectively measured (Lifecorder EX)
Patel et al [26][b]	Adults, n=200 (mean age 55.9 years, SD 9.9 years; 44% male participants; BMI 27.1)	Participants were entered into a game with their family in teams and had to complete weekly goal targets to win points and badges; duration: 12 weeks	Behavioral economics	Daily step count objectively measured (Withings wrist-worn device)
Patel et al [27][b]	Adults with overweight and obesity, n=602 (mean age 38.7 years, SD 10.4 years; 69% male participants; BMI 29.6)	Participants had to complete weekly goal targets to win points and levels. There were 3 versions of the intervention: support, collaboration, and competition; duration: 24 weeks	Behavioral economics	Daily step count objectively measured (Withings wrist-worn device)
Paul et al [28][b]	Patients who survived stroke, n=23 (mean age 55.8 years, SD 10.7 years; 48% male participants; BMI 24.5)	In the "STARFISH" app, participants had to complete their PA objectives to improve their avatar; duration: 6 weeks	Control theory and Michie taxonomy of behavior change	Daily step count, sedentary time, and walking time (min/week) objectively measured (ActivPAL)
Thorsteinsen et al [29][b]	Adults, n=21 (mean age 55.3 years, SD 11.2 years; 52% male participants)	The intervention "Lifestyle Tool" consisted of a rule-based website designed to help people plan and monitor their PA. The tool incorporated social and individual gaming components to increase motivation and engagement; duration: 12 weeks	—	Self-reported weekly activity minutes (daily report form)
Zuckerman and Gal-Oz [30][b]	Students, n=59 (mean age 23.4 years, SD 1.4 years; 25% male participants)	"StepByStep" was an accelerometer-based mobile app with virtual rewards and social comparison intended to motivate people to incorporate more walking into their daily routine; duration: 1.5 week	SDT	Walking time (min/day) objectively measured (smartphone accelerometer)

[a]PA: physical activity.

[b]The studies included in the meta-analysis.

[c]SDT: self-determination theory.

[d]MVPA: moderate to vigorous physical activity.

[e]SCT: sociocognitive theory.

[f]TPB: theory of planned behavior.

[g]VPA: vigorous physical activity.

[h]MPA: moderate physical activity.

[i]IPAQ: International Physical Activity Questionnaire.

[j]LPA: light physical activity.

[k]PAQ-A: Physical Activity Questionnaire for Adolescents.

[l]No theory mentioned.

Studies comparing gamified interventions with active control groups used a similar intervention without game elements (ie, an equivalent nongamified app [30-32] or a self-monitoring intervention with wearables or activity monitors [25-27,34,36]).

In most studies, the interventions were based on theoretical models. A total of 6 studies [21,30,31,33,51,52] were based on the self-determination theory [54]; 5 [22-24,32,34] on sociocognitive models (ie, the transtheoretical model [55], the social cognitive theory [56], the theory of planned behavior [57], and the health action process approach [58]); and 3 [26,27,36] on behavioral economics models.

Outcomes measured in trials were diverse: they used either total PA duration or MVPA duration, SB duration, daily step count, walking duration, or active minute count. A total of 13 experiments measured PA objectively using devices such as triaxial accelerometers (n=7; [21,28,31-33,51,52]), wearable

devices for the general population (ie, Fitbit, Garmin, Polar, and Withings monitors; n=6; [23,26,27,34-36]), pedometers (n=1; [25]), or smartphones (n=1; [30]), and 5 assessed PA with self-reported measures (International Physical Activity Questionnaire [22], Physical Activity Questionnaire for Adolescents [31], Active Australian Survey [24,32], or other [29]). A total of 6 trials [21,24,26,27,32,36] completed a follow-up assessment from 12 to 30 weeks (mean 14.4 weeks) after the end of the intervention.

A total of 2 studies [51,52] were excluded from the meta-analysis and were only integrated in the qualitative review because we were unable to extract their results.

Risk of Bias

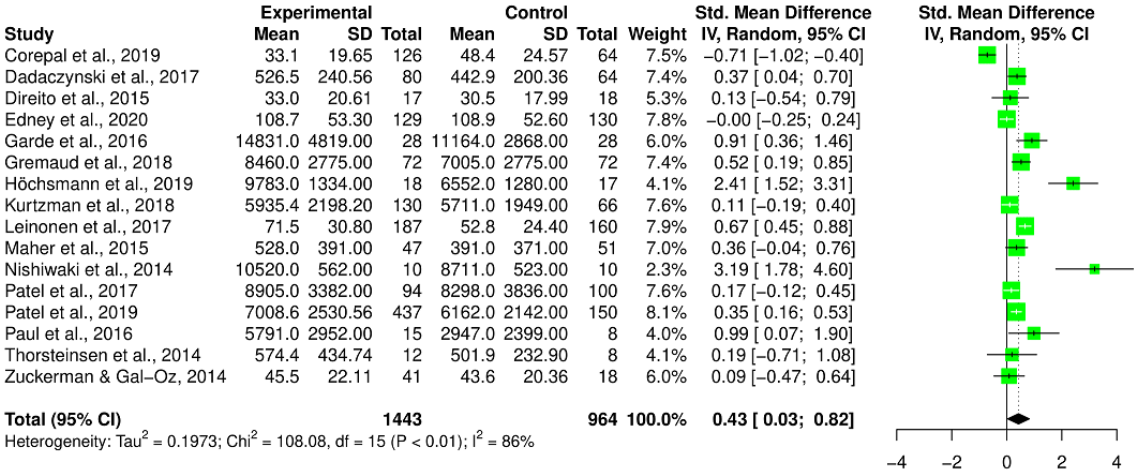
The 2 authors extracted the risk of bias data with a κ coefficient of 0.79, which is synonymous with excellent agreement [53]. Multimedia Appendices 2 [21-36] and 3 present the authors' judgments about each risk of bias item presented as percentages across all included studies in the meta-analysis and an overview of the different biases for each study. Overall, 1 study [28] was rated as high risk for sequence generation because assignments were based on recruitment order. Therefore, this study was also at a high risk for allocation concealment. A total of 3 studies [24,29,30] were at high risk of bias for the *blinding of outcome assessment* item because they measured PA using only self-reports. In total, 5 studies [23,29,30,49,52] were at high risk of bias for the *incomplete outcome data* item because they reported high dropout rates and did not include intention-to-threat analyses and 5 studies [25,28,29,31,52] were rated at unclear risk for the *selective outcome reporting* item because they had not been preregistered or published in a protocol-study. Finally, 2 studies had other high risks of bias. The first one [25] was a crossover trial conducted without a washout condition, and the other one [21] was a cluster randomized trial with no control of clustering, no consideration of the clustering in the statistical analysis, and no test of baseline differences among groups.

Summary Effect

Overall, the SMD after the intervention for all PA outcomes (MVPA, daily step count, number of active minutes, and walking time) was a Hedges *g* of 0.43 (95% CI 0.03-0.82; I^2=86%; Figure 2), representing a statistically significant small to medium effect. Similarly, we found a statistically significant SMD effect of a Hedges *g* of 0.38 (95% CI 0.07-0.69; I^2=79%) for pre–post change scores.

Only 3 studies [23,28,31] assessed sedentary time as an outcome. Owing to this small sample size, the meta-analysis was not performed on this outcome.

Figure 2. Forest plot for the effect of gamification versus control on postintervention physical activity outcomes (moderate to vigorous physical activity, daily step count, number of active minutes, and walking time). Tau-square, chi-square, and I^2 measures of between-study heterogeneity [21-36]. IV: inverse variance.



Outliers and Influential Analyses

In the first analysis, substantial statistical heterogeneity was observed. To address between-study heterogeneity, we first looked for the presence of outliers. A total of 3 studies were considered as outliers [21,25,35], and after removing them, we still obtained a significant effect of a Hedges *g* of 0.34 (95% CI 0.17-0.51) with moderate heterogeneity (I^2=58%).

Leave-one-out analyses showed that sequential removal of each study did not have an important impact on the general effect size, with effect sizes ranging from a Hedges *g* of 0.33 (95% CI 0.00-0.66; I^2=84%) to a Hedges *g* of 0.48 (95% CI 0.13-0.83; I^2=78%; Figure S2 in Multimedia Appendix 1).

The Baujat plot (Figure S3 in Multimedia Appendix 1) shows that 4 studies explained more heterogeneity than the others, more specifically, the study by Corepal et al [21] with a heterogeneity contribution of 40.13 and an effect size influence of 3.27.

Therefore, we excluded studies with a high or unclear risk of bias. After doing so, the effect was not significant (Hedges *g*=0.33, 95% CI –0.16 to 0.81; I^2=78%).

The inclusion of crossover and cluster randomized trials in the meta-analysis may lead to biases. Thus, we excluded these designs from the analysis and obtained a statistically significant effect of *g*=0.49 (95% CI 0.05-0.92; I^2=67%).

Finally, we decided to exclude articles by Corepal et al [21] and Nishiwaki et al [25] in the sensitivity analyses (which have been repeated afterward for each analysis) considering their influence on the pooled result, their contribution to overall heterogeneity, and their huge risk of bias (no control of clustering, no statistical consideration of clustering, and no test of baseline differences among groups in the study by Corepal et al [21] and no washout period and very low power for the study by Nishiwaki et al [25]). After doing so, we obtained a statistically significant effect of a Hedges *g* of 0.42 (95% CI 0.14-0.69; I^2=74%).

Subgroup Analyses

We found no statistical differences in the effects between studies with participants with chronic diseases or healthy participants (Cochran *Q*=0.73; *P*=.39), between adults and adolescents (Cochran *Q*=0.26; *P*=.61), between studies with objective (devices) or self-reported PA outcomes (Cochran *Q*=0.23; *P*=.63), between active or inactive control groups (Cochran *Q*=0.01; *P*=.92), and between short- and long-term interventions (less or more than 12 weeks; Cochran *Q*=0.60; *P*=.44).

When performing the sensitivity analysis, there was a statistically significant effect of intervention on PA in adults (Hedges *g*=0.36, 95% CI 0.03-0.69; I^2=71%; Figure S4 in Multimedia Appendix 1), on healthy people (Hedges *g*=0.35, 95% CI 0.15-0.55; I^2=63%; Figure S5 in Multimedia Appendix 1), when the PA measure was objective (Hedges *g*=0.45, 95% CI 0.08-0.82; I^2=80%; Figure S6 in Multimedia Appendix 1), when the PA measure was self-reported (Hedges *g*=0.24, 95% CI 0.08-0.39; I^2=0%; Figure S6 in Multimedia Appendix 1), and for short interventions of <12 weeks (equivalent to a 3-month program; Hedges *g*=0.44, 95% CI 0.19-0.69; I^2=16%; Figure S7 in Multimedia Appendix 1).

Moreover, subgroup analyses allowed us to examine the effect of gamified interventions when compared with inactive control groups and active control groups. After sensitivity analyses, we found a statistically significant effect of gamified interventions, both when compared with inactive control groups (Hedges *g*=0.58, 95% CI 0.08-1.07; I^2=81%; Figure S8 in Multimedia Appendix 1) and when compared with active control groups (Hedges *g*=0.23, 95% CI 0.05-0.41; I^2=37%; Figure S8 in Multimedia Appendix 1).

Meta-Regressions

The age of participants (β=.01, 95% CI –0.02 to 0.04; *P*=.39), their gender (β=.01, 95% CI –0.01 to 0.02; *P*=.47), their BMI (β=.04, 95% CI –0.16 to 0.09; *P*=.53), the duration of the intervention (β=-.01, 95% CI –0.06 to 0.04; *P*=.74), or the number of game mechanics included in the intervention (β=.01, 95% CI –0.17 to 0.19; *P*=.91) were not statistically significantly associated with an increase in PA.

Lack of data precluded further meta-regressions, such as comparisons of leisure PA, or test of moderators, such as the impact of social incentives or the theoretical model used to develop the intervention.

Publication Bias

First, an inspection of the funnel plot showed that the effect sizes of individual studies were relatively symmetrically distributed around the pooled effect size. This observation was supported by the Egger test of the intercept, which indicated no asymmetry in the funnel plot (b_0=1.38, 95% CI –0.83 to 4.77; *P*=.19). We then applied a bias-correction technique, the trim-and-fill method, which indicated that 3 studies were missing at the bottom left of the funnel plot to obtain a full symmetry. After imputing the effect sizes corresponding of these missing studies to obtain a totally symmetrical funnel plot (Figure 3), the bias corrected summary effect was of a Hedges *g* of 0.24 (95% CI –0.24 to 0.73).

Finally, the sunset funnel plot (Figure 4) showed significant differences in power among studies, with some characterized by very low statistical power (7 studies under 45% power and 4 studies under 18%). The median power of all the tests was 63%.

Figure 3. Funnel plot after trim-and-fill bias correction. A filled circle represents an included study, and an empty circle represents a missing study.

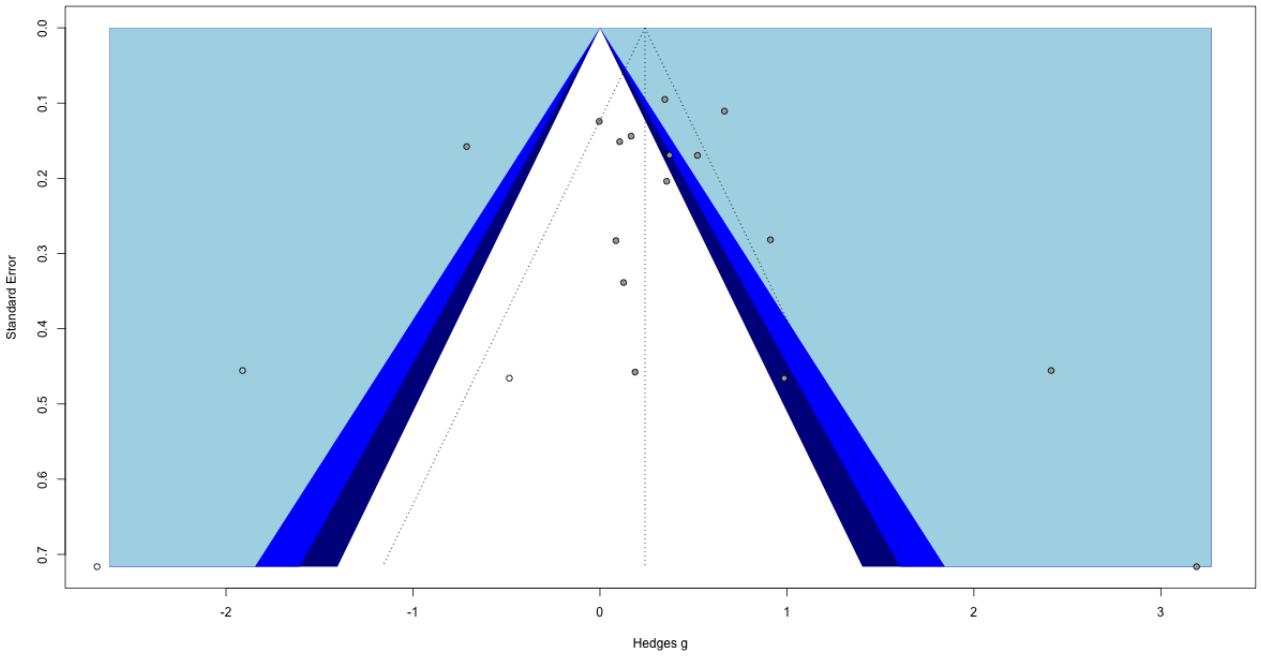
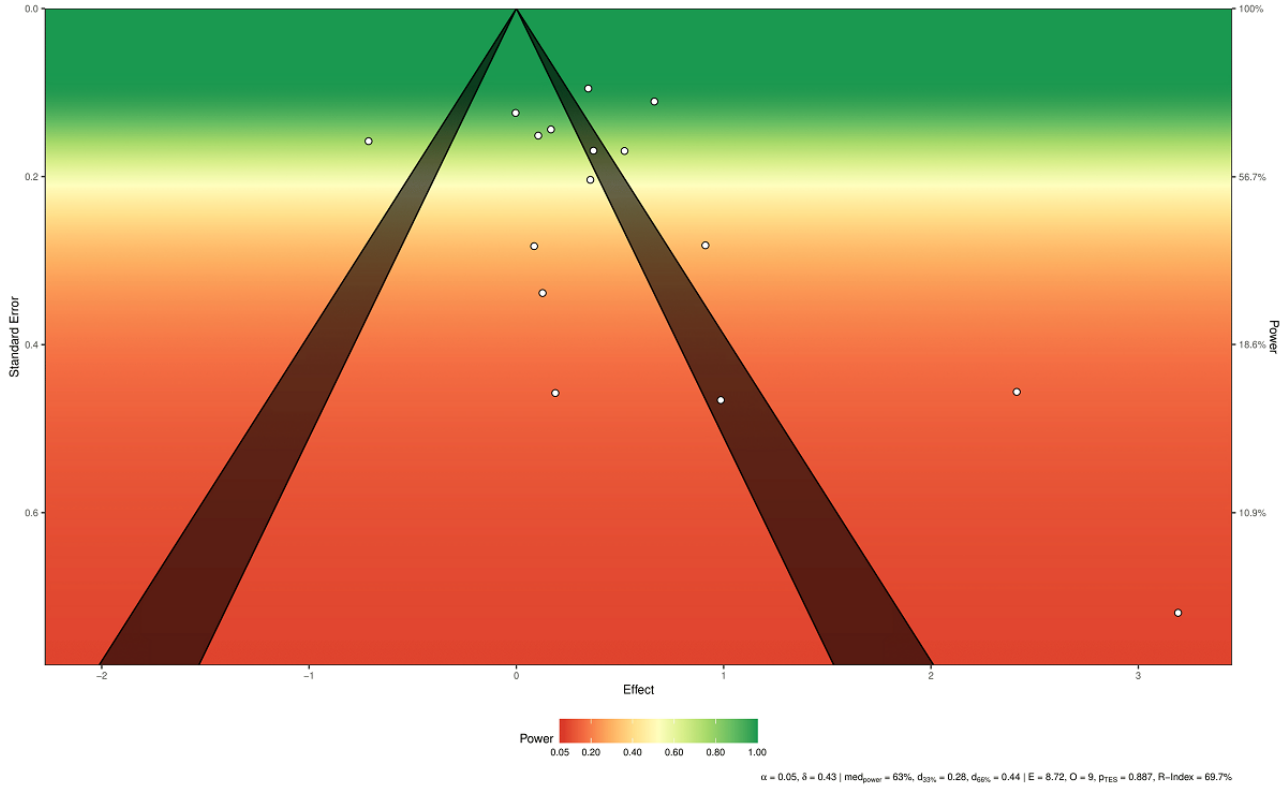


Figure 4. Power-enhanced funnel plot. White circles represent included studies. δ: true effect size; medpower: the median power of all tests; d33%: effect size needed for achieving 33% of median power; d66%: effect size needed for achieving 66% of median power; E: expected number of positive studies; O: observed number of positive studies; pTES: test of excess significance *P* value.

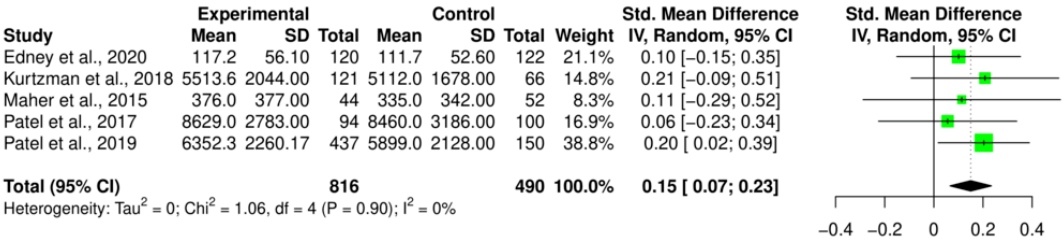


Secondary Analyses

Follow-up

There was no statistically significant effect of gamified interventions on total PA (MVPA and daily step count) after follow-up periods with an SMD of a Hedges *g* of 0.09 (95% CI –0.07 to 0.26; I^2=21%). When we performed the sensitivity analysis, gamification significantly increased PA (MVPA and daily step count) at follow-up (from 12 to 24 weeks after the end of the intervention; *g*=0.15, 95% CI 0.07-0.23; I^2=0%; Figure 5).

Figure 5. Forest plot for the effect of gamification versus control on PA outcomes (moderate to vigorous physical activity and daily step count) after a follow-up period (from 12 to 24 weeks after the end of the intervention). Tau-square, chi-square, and I^2 measures of between-study heterogeneity [24,26,27,32,36]. IV: inverse variance.



Steps

We found no statistically significant effect of gamified interventions on step outcomes with an SMD of a Hedges *g* of 0.53 (95% CI –0.09 to 1.15; I^2=89%), but a significant improvement in the number of daily steps with an MD of +1420.57 steps per day (95% CI 435.41-2405.73; I^2=95%) was observed. When excluding the 2 studies in the sensitivity analysis, we obtained a statistically significant effect of gamification on daily steps of a Hedges *g* of 0.49 (95% CI 0.05-0.93; I^2=75%; Figure 6) and a statistically significant MD of +1609.56 steps per day (95% CI 372.39-2846.73; I^2=86%; Figure 7).

Figure 6. Forest plot for the effect of gamification versus control on steps outcomes (daily step count and walking time). Tau-square, chi-square, and I^2 measures of between-study heterogeneity [22,26-28,30,33-36]. IV: inverse variance.

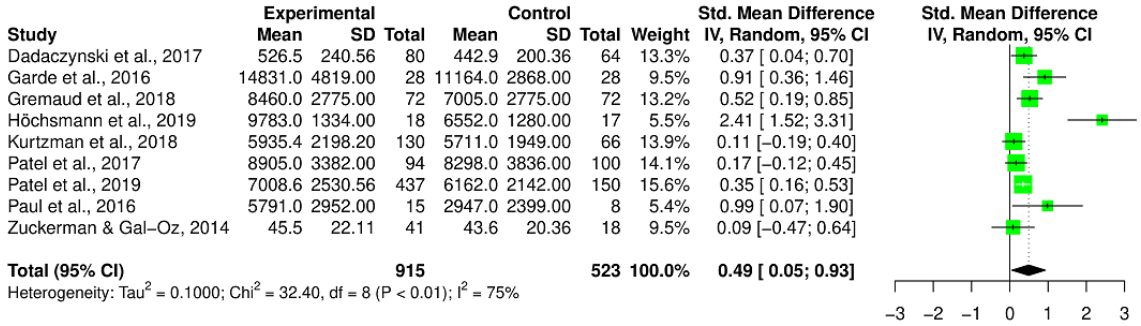
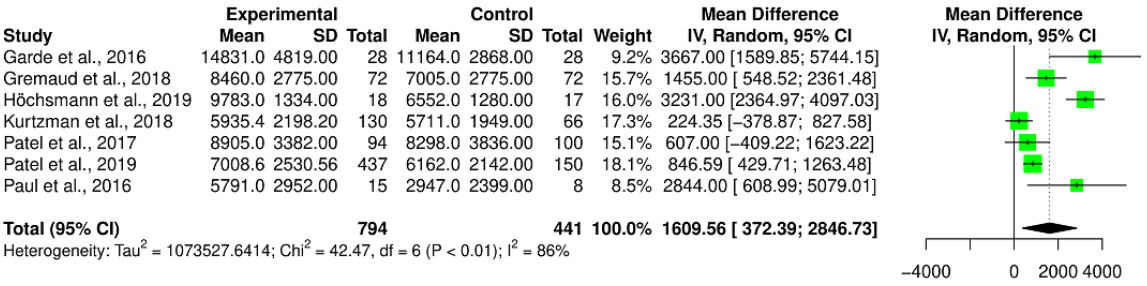


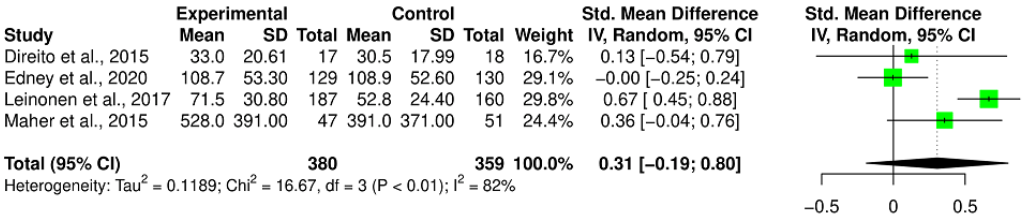
Figure 7. Forest plot for the mean difference of daily steps between gamification and control. Tau-square, chi-square, and I^2 measures of between-study heterogeneity [26-28,33-36]. IV: inverse variance.



Moderate to Vigorous PA

There was no statistically significant effect of gamification on MVPA with an SMD of a Hedges *g* of 0.09 (95% CI –0.57 to 0.74; I^2=93%). There was no statistically significant effect of a Hedges *g* of 0.31 (95% CI –0.19 to 0.80; I^2=82%) in the sensitivity analysis (Figure 8).

Figure 8. Forest plot for the effect of gamification versus control on moderate to vigorous physical activity. Tau-square, chi-square, and I^2 measures of between-study heterogeneity [23,24,31,32]. IV: inverse variance.



Summary of Findings (Grading of Recommendations Assessment, Development, and Evaluation)

The quality of evidence (grading of recommendations assessment, development, and evaluation, [GRADE]) in the included studies after sensitivity analyses for short-term PA, long-term PA, MVPA, steps, and daily steps was scored from high to low (Table 2). The quality was downgraded for some outcomes because of high heterogeneity, high risk of bias, or imprecision owing to large CIs. Summaries of the various meta-analysis conducted in this review both on postintervention scores and pre–post intervention change scores are presented in Multimedia Appendix 1 (Figures S9-S11).

Table 2. Summary of findings.

Outcome	Number of participants (number of studies)	Standardized mean difference or mean difference (95% CI)	Quality of evidence (grading of recommendations assessment, development, and evaluation)
General PA[a]	2197 (14)	0.42 (0.14 to 0.69)	Low[b,c,d]
General PA (in comparison with active control groups)	1485 (7)	0.23 (0.05 to 0.41)	High
Long-term PA (follow-up)	1306 (5)	0.15 (0.07 to 0.23)	High
MVPA[e]	739 (4)	0.31 (–0.19 to 0.80)	Low[b,c,d]
Steps	1438 (9)	0.49 (0.05 to 0.93)	Low[b,c,d]
Daily steps	1235 (7)	1609.56 (372.39 to 2846.73)	Moderate[b,d]

[a]PA: physical activity.

[b]Downgraded because of high heterogeneity.

[c]Downgraded because of risks of bias.

[d]Downgraded because of imprecision (large CIs).

[e]MVPA: moderate to vigorous physical activity.

Discussion

Principal Findings

Summary Effect

This meta-analysis of RCTs, including 16 studies and 2407 participants, revealed a statistically significant effect of gamified interventions, on average, of 12 weeks on total PA (Hedges *g*=0.42, 95% CI 0.14-0.69 after sensitivity analyses). This effect was small to medium, suggesting the effectiveness of gamified interventions in promoting PA in both healthy participants and participants with chronic diseases. This significant effect was robust, as it persisted even after the different influence analyses were performed. Moreover, the effect was statistically significant both for objective measures of PA (Hedges *g*=0.45, 95% CI 0.08-0.82) and self-reported measures (Hedges *g*=0.24, 95% CI 0.08-0.39) after sensitivity analyses. Unsurprisingly, subgroup analyses revealed after sensitivity analyses that the effect of gamified interventions is greater when compared with inactive control groups (such as waiting lists) than when compared with active control groups benefiting from a nongamified intervention (Hedges *g*=0.58, 95% CI 0.08-1.07 vs Hedges *g*=0.23, 95% CI 0.05-0.41). Nevertheless, these effects were both statistically significant. This suggests that gamified interventions are not only efficient in changing behavior but also, to a lesser extent, effective compared with equivalent nongamified PA interventions (such as smartphone apps or self-monitoring interventions). These results are important considering the assets of gamification, which has the advantages of (1) reorganizing existing activity rather than adding additional demands to people's lives [13], (2) being easily implemented in natural contexts, and (3) having a broad accessibility through technology and advancing sensors, permitting to address a large population.

Long-term Effect

When we analyzed the long-term effect of these interventions based on the follow-up measures of PA, carried out from 12 to 24 weeks (mean 14.4 weeks) after the end of the intervention, we found a statistically significant very small to small effect size of a Hedges *g* of 0.15 (95% CI 0.07-0.23) after sensitivity analyses. These results indicate that the effect of gamification persists after the end of the program, suggesting that it is not just a novelty effect due to the playful nature of gamification. However, this long-term effect was weaker and decreased with time after the end of the intervention.

Generalizability of Gamified Interventions

The absence of subgroup differences or effects of age, gender, and BMI on the pooled effect suggests a good generalizability of gamified interventions, which can be used for several types of populations. Thus, gamification may not only be efficient in young healthy individuals but can also target any kind of population regardless of their age or health status.

In sum, gamified interventions appear as an efficient tool to improve the PA of various populations, with moderate superiority over other similar interventions, such as mobile health monitoring apps, and a moderate sustainability of the effect after the intervention. Nevertheless, if many PA interventions increase PA levels in the short term, translating these temporary changes into long-term PA participation continues to be a challenge for PA research [59]. With that in mind, the potential of gamification for PA increases in the long

term, even minimal, is particularly important and promising in the area of PA interventions.

Additional Findings

Effect of Gamification on the Step Count

If the overall effect of gamified interventions on PA is positive, they increase the step count more than MVPA. Indeed, after sensitivity analyses, on the one hand, the meta-analyses revealed a statistically significant effect (Hedges *g*=0.49, 95% CI 0.05-0.93) of gamification for steps outcomes, with a statistically significant improvement of 1609.56 steps per day (95% CI 372.39-2846.73) for participants benefiting from gamified intervention versus those in the control group. On the other hand, no statistically significant effect of gamified interventions on MVPA was found (Hedges *g*=0.31, 95% CI –0.19 to 0.80). This can be explained by the game metrics and mechanics of the interventions included in the review, which are mainly focused on the step count of participants. Few interventions directly targeted MVPA. In the included studies, only 2 interventions [31,35] integrated multi-PA intensity goals and mechanics, notably with physical exercises or running sessions in the game. In other words, participants played most of the time with their number of steps and had to generally walk more to make points and play the game. This results in an increase in walking time but not necessarily in more intense PA.

These findings are interesting considering the potential health benefits of increasing the number of daily steps by 1600 because of gamification. Indeed, previous work showed that walking was statistically associated with decline in all-cause and cardiovascular mortality [60-63] and an improvement in body composition [64]. Moreover, Oja et al [62] suggested that any walking exposure is beneficial for cardiovascular health, endorsing the idea that the most important is more global PA regardless of the intensity [7,65] even when this activity only includes walking [60]. In comparison, a previous study evaluating the effectiveness of activity trackers with and without incentives to increase PA [66] showed a significant improvement of 1050 daily steps for the cash incentive intervention versus the control intervention (95% CI 600-1490) but no statistically significant difference for the Fitbit-only group (340 daily steps, 95% CI –100 to 790]). In light of these results, gamified interventions appear as an added value compared with current interventions. Considering that 40% of the volunteers in this study abandoned their Fitbit monitor within 6 months, gamification is also a way to keep participants involved and motivated within the intervention.

Duration of Intervention

Our meta-regression analysis did not find an association between the observed effect of gamification on PA and intervention length. However, although no statistically significant effect of gamification for an intervention length of ≥12 weeks was found (Hedges *g*=0.41, 95% CI –0.19 to 1.01), the meta-analysis revealed a statistically significant effect of gamified interventions of <12 weeks on global PA (Hedges *g*=0.44, 95% CI 0.19-0.69). According to a previous meta-analysis that reported significant positive effects of smartphone apps on PA only when used over a short-term period of <3 months [67], these results suggest that a condensed intervention could benefit more than a longer one, which could become redundant, boring, and exhausting for participants in the long run.

Statistical Heterogeneity

The meta-analysis also revealed considerable statistical heterogeneity. This heterogeneity may be explained by differences in study quality, diversity of designs, and variations in study populations. Despite several subgroup analyses, we cannot rule out that these subgroups and characteristics may not explain all the variance of the interventions. Indeed, demographic data often do not fully explain the differences in the effectiveness of interventions [68,69], and more precise sociopsychological variables such as personality traits or motivational factors could explain this poorly understood significant heterogeneity. Moreover, the risk of bias analysis and the sunset funnel plot showed substantial differences in the quality of the included studies that can influence the heterogeneity. Finally, various trial designs were used (ie, parallel RCT, cluster RCT, and crossover RCT) that can also contribute to the overall statistical heterogeneity.

Better Understanding Gamification Mechanisms

This meta-analysis is informative regarding the effectiveness of gamified interventions. In view of the observed heterogeneity, the next step will be to investigate its causes from an interventional and theoretical perspective. Gamified interventions involve multiple interacting elements, and it is crucial to estimate the weight of each element in the behavior change process and how they interact with each other. Is it game mechanics, the implementation of behavior change techniques, or the presence of social interactions that make gamification effective? Unfortunately, the small number of studies included in the meta-analysis impeded us from conducting in-depth moderation analyses to answer this question. To better understand these relations, it is essential that both the development and assessment of gamified interventions be central, transparent, evidence-based, context-aware, and research-oriented [70]. Moreover, if theoretical psychological models are often mentioned in the introduction of articles included in the review (Table 1), few have investigated the psychological mechanisms of their interventions in the field [22,31,71]. Future studies should explicitly discuss motivational theory and systematically test the effect of gamification on psychological outcomes known to be involved in behavior change (eg, self-efficacy, attitudes, and intention) to better understand its mechanisms. The consideration of personality traits and psychological variables to determine behavioral phenotypes [68] is a promising way to evaluate participant's responses to the interventions.

Perspectives for Future Research and Implications for Practice

The findings from this meta-analysis allow us to draw and discuss future work concerning the gamification of PA and SB. First, future trials should be conducted with more adequately powered sample sizes and should be strictly multiple arms–RCTs to isolate the effects of gamification elements and better understand gamification mechanisms. Second, the

long-term effects are currently the main challenge of health interventions. Thus, it is essential to investigate the evolution of the effects of gamified interventions over time. Therefore, there is a need for more long-term follow-up measurements. In addition, the potentialities of digital technologies and their capacity to collect a large amount of real-world data could be used to assess the evolution patterns of the effect, allowing the detailed identification of its sustainability and evolvement or even make forecasts via time series analyses. Third, to our knowledge, only 1 team of researchers worked on a gamified intervention targeting SB by introducing sedentary breaks as a gaming part [72]. Following this line of research, it could be interesting to develop gamified interventions affording participants to take more sedentary breaks. Finally, the cost-effectiveness ratio of gamified interventions may be better than that of many current interventions, considering the ease of implementation and generalizability of gamification. However, this assertion will have to be tested in future trials, including economic analyses.

In light of our results, gamified interventions appear to be a promising avenue to promote PA in different populations both in prevention in healthy people and in the treatment of chronic diseases. Gamified interventions have many benefits for participants with chronic diseases, such as empowerment of participants by improving their self-management skills, an effect across broad audiences enabling to target different types of pathologies, and an everyday life fit and easy implementation. Similar to other digital health processes, gamification makes it possible to address more patients, especially those who are isolated from health care facilities. Importantly, gamified interventions are especially pertinent during a health pandemic, such as the COVID-19 outbreak, in which PAs and social interactions are restricted because of lockdown or teleworking and where structured PA possibilities are limited both indoors and outdoors. Gamifying walking and daily activities is, in this context, a great way to improve PA and limit SB of individuals in addition to providing social interaction among players. In the meantime, the face-to-face management of chronic diseases is usually suspended during the pandemic, which underlines the importance of offering remote supervision of PA.

Nevertheless, in view of the weaker postintervention effect, this study suggests that a one-shot intervention is not sufficient. A more interesting design would be to address multiple *gamification doses* during or after the course of treatment to obtain a sustainable implementation of the PA behavior. This configuration would also provide an ideal duration of intervention to avoid exhausting the participants with gamified interventions for >12 weeks.

Strengths and Limitations

To our knowledge, this is the first meta-analysis to quantitatively evaluate the effects of gamification on PA. This review has several other strengths. First, we conducted a comprehensive search strategy using multiple databases in collaboration with an academic librarian. Second, all stages of the review (screening and data extraction) were independently realized by 2 reviewers. Finally, various novel publication bias analyses and influence analyses were conducted in parallel with different subgroup analyses and meta-regressions.

However, some limitations of this work must be mentioned. Overall, the meta-analysis included a small number of studies, and some articles were feasibility or pilot trials. Therefore, several trials included small sample sizes and were highly underpowered. Some studies were conducted with a high risk of bias. One of the main limitations of this work is the impossibility of demonstrating that the effect of gamified interventions is led by gamification itself given the lack of research examining this question. Finally, in the main analysis, we included diverse PA outcomes evaluating similar constructs but which are slightly different in practice. Moreover, not all included outcomes were objectively measured. As the field matures and new trials are published, an update of this work will be important to confirm these preliminary results.

Conclusions

To conclude, gamified interventions appear to be a promising avenue for promoting PA in various populations. Influencing primarily the number of daily steps of the participants, gamification is an interesting way to improve daily PA and appears more efficient than equivalent nongamified interventions, such as mobile health apps. However, if the effect of gamification persists during follow-up, suggesting that gamified interventions are more than a novelty effect, this effect decreases with time with a smaller long-term effect. The integration of gamification in more global health care interventions could be a way to address this limited sustainability. Future rigorous trials are required to explore these perspectives.

Acknowledgments

The authors would like to thank the librarian at Clermont Auvergne University (Nathalie Pinol) for assistance in developing the search equations. The work of AM is supported by a grant from the French National Association for Research and Technology (Cifre PhD thesis) and by the company Kiplin. None of the sponsors were involved in any other aspect of the project, such as the design of the project's protocol and analysis plan, collection, and analyses. The funders had no input on the interpretation or publication of the study results. The authors also want to thank Challenge 3 I-SITE Clermont Auvergne Project 20-25 for their grant.

Authors' Contributions

AM designed the study protocol and search strategy in consultation with other authors. AM and AC conducted the screening, data extraction, and risk of bias assessment. AM and BP developed the meta-analytic procedure, and AM performed the statistical

analysis. The first draft of the manuscript was written by AM, and all authors commented on the previous versions of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[PDF File (Adobe PDF File), 999 KB - jmir_v24i1e26779_app1.pdf]

Multimedia Appendix 2
Risk of bias summary for studies included in the meta-analysis.
[PDF File (Adobe PDF File), 301 KB - jmir_v24i1e26779_app2.pdf]

Multimedia Appendix 3
Risk of bias graph: review authors' judgments about each risk of bias item presented as percentages across all included studies.
[PDF File (Adobe PDF File), 59 KB - jmir_v24i1e26779_app3.pdf]

References

1. Hallal PC, Andersen LB, Bull FC, Guthold R, Haskell W, Ekelund U, Lancet Physical Activity Series Working Group. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet 2012 Jul 21;380(9838):247-257. [doi: 10.1016/S0140-6736(12)60646-1] [Medline: 22818937]
2. Lee I, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, Lancet Physical Activity Series Working Group. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet 2012 Jul 21;380(9838):219-229 [FREE Full text] [doi: 10.1016/S0140-6736(12)61031-9] [Medline: 22818936]
3. Manson JE, Greenland P, LaCroix AZ, Stefanick ML, Mouton CP, Oberman A, et al. Walking compared with vigorous exercise for the prevention of cardiovascular events in women. N Engl J Med 2002 Sep 05;347(10):716-725. [doi: 10.1056/NEJMoa021067] [Medline: 12213942]
4. Wolin KY, Yan Y, Colditz GA, Lee I. Physical activity and colon cancer prevention: a meta-analysis. Br J Cancer 2009 Feb 24;100(4):611-616 [FREE Full text] [doi: 10.1038/sj.bjc.6604917] [Medline: 19209175]
5. Duclos M, Oppert J, Verges B, Coliche V, Gautier J, Guezennec Y, SFD Diabetes and Physical Activity Working Group. Physical activity and type 2 diabetes. Recommandations of the SFD (Francophone Diabetes Society) diabetes and physical activity working group. Diabetes Metab 2013 May;39(3):205-216. [doi: 10.1016/j.diabet.2013.03.005] [Medline: 23643351]
6. Pedersen BK, Saltin B. Exercise as medicine - evidence for prescribing exercise as therapy in 26 different chronic diseases. Scand J Med Sci Sports 2015 Dec;25 Suppl 3:1-72. [doi: 10.1111/sms.12581] [Medline: 26606383]
7. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. Br Med J 2019 Aug 21;366:l4570 [FREE Full text] [doi: 10.1136/bmj.l4570] [Medline: 31434697]
8. Chau JY, Grunseit AC, Chey T, Stamatakis E, Brown WJ, Matthews CE, et al. Daily sitting time and all-cause mortality: a meta-analysis. PLoS One 2013 Nov 13;8(11):e80000 [FREE Full text] [doi: 10.1371/journal.pone.0080000] [Medline: 24236168]
9. Ekelund U, Tarp J, Fagerland M, Johannessen J, Hansen B, Jefferis B, et al. Joint associations of accelero-meter measured physical activity and sedentary time with all-cause mortality: a harmonised meta-analysis in more than 44 000 middle-aged and older individuals. Br J Sports Med 2020 Dec;54(24):1499-1506 [FREE Full text] [doi: 10.1136/bjsports-2020-103270] [Medline: 33239356]
10. Deterding S, Dixon D, Khaled R, Nacke L. From game design elements to gamefulness: defining "gamification". In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments. 2011 Presented at: MindTrek '11: Academic MindTrek 2011; September 28 - 30, 2011; Tampere Finland p. 9-15. [doi: 10.1145/2181037.2181040]
11. Cugelman B. Gamification: what it is and why it matters to digital health behavior change developers. JMIR Serious Games 2013 Dec;1(1):e3 [FREE Full text] [doi: 10.2196/games.3139] [Medline: 25658754]
12. Hamari J, Koivisto J, Sarsa H. Does Gamification Work? -- A Literature Review of Empirical Studies on Gamification. In: Proceedings of the 47th Hawaii International Conference on System Sciences. 2014 Presented at: 47th Hawaii International Conference on System Sciences; January 6-9, 2014; Waikoloa, HI, USA. [doi: 10.1109/HICSS.2014.377]
13. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]

14. Munson S, Poole E, Perry DB, Peyton T. Gamification and health. In: Walz SP, Deterding S, editors. The Gameful World: Approaches, Issues, Applications. Cambridge, London: MIT Press; 2015:597-624.
15. Koivisto J, Hamari J. Gamification of physical activity: a systematic literature review of comparison studies. In: Proceedings of the 3rd International GamiFIN Conference. 2019 Presented at: 3rd International GamiFIN Conference, GamiFIN 2019; April 8-10, 2019; Levi, Finland p. 106-117 URL: https://researchportal.tuni.fi/en/publications/gamification-of-physical-activity-a-systematic-literature-review-
16. Seaborn K, Fels DI. Gamification in theory and action: a survey. Int J Hum-Comput Stud 2015 Feb;74:14-31. [doi: 10.1016/j.ijhcs.2014.09.006]
17. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA Statement. Open Med 2009;3(3):123-130 [FREE Full text] [Medline: 21603045]
18. Quintana DS. From pre-registration to publication: a non-technical primer for conducting a meta-analysis to synthesize correlational data. Front Psychol 2015 Oct 08;6:1549 [FREE Full text] [doi: 10.3389/fpsyg.2015.01549] [Medline: 26500598]
19. Mazéas A, Duclos M, Pereira B, Chalabaev A. Does gamification improve physical activity? A systematic review and meta-analysis. Center for Open Science (OSF) 2020. [doi: 10.17605/OSF.IO/X2483]
20. Moreau D, Gamble B. Conducting a meta-analysis in the age of open science: tools, tips, and practical recommendations. Psychol Methods 2020 Sep 10. [doi: 10.1037/met0000351] [Medline: 32914999]
21. Corepal R, Best P, O'Neill R, Kee F, Badham J, Dunne L, et al. A feasibility study of 'The StepSmart Challenge' to promote physical activity in adolescents. Pilot Feasibility Stud 2019 Nov;5:132 [FREE Full text] [doi: 10.1186/s40814-019-0523-5] [Medline: 31832227]
22. Dadaczynski K, Schiemann S, Backhaus O. Promoting physical activity in worksite settings: results of a German pilot study of the online intervention Healingo fit. BMC Public Health 2017 Sep 08;17(1):696 [FREE Full text] [doi: 10.1186/s12889-017-4697-6] [Medline: 28886734]
23. Leinonen A, Pyky R, Ahola R, Kangas M, Siirtola P, Luoto T, et al. Feasibility of gamified mobile service aimed at physical activation in young men: population-based randomized controlled study (MOPO). JMIR Mhealth Uhealth 2017 Oct 10;5(10):e146 [FREE Full text] [doi: 10.2196/mhealth.6675] [Medline: 29017991]
24. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De BI, Thomas S, et al. A web-based, social networking physical activity intervention for insufficiently active adults delivered via facebook app: randomized controlled trial. J Med Internet Res 2015;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
25. Nishiwaki M, Kuriyama A, Ikegami Y, Nakashima N, Matsumoto N. A pilot crossover study: effects of an intervention using an activity monitor with computerized game functions on physical activity and body composition. J Physiol Anthropol 2014 Dec 02;33:35 [FREE Full text] [doi: 10.1186/1880-6805-33-35] [Medline: 25466450]
26. Patel MS, Benjamin EJ, Volpp KG, Fox CS, Small DS, Massaro JM, et al. Effect of a game-based intervention designed to enhance social incentives to increase physical activity among families: the BE FIT randomized clinical trial. JAMA Intern Med 2017 Nov 01;177(11):1586-1593 [FREE Full text] [doi: 10.1001/jamainternmed.2017.3458] [Medline: 28973115]
27. Patel MS, Small DS, Harrison JD, Fortunato MP, Oon AL, Rareshide CA, et al. Effectiveness of behaviorally designed gamification interventions with social incentives for increasing physical activity among overweight and obese adults across the united states: the STEP UP randomized clinical trial. JAMA Intern Med 2019 Sep 09:1-9 [FREE Full text] [doi: 10.1001/jamainternmed.2019.3505] [Medline: 31498375]
28. Paul L, Wyke S, Brewster S, Sattar N, Gill JMR, Alexander G, et al. Increasing physical activity in stroke survivors using STARFISH, an interactive mobile phone application: a pilot study. Top Stroke Rehabil 2016 Jun;23(3):170-177. [doi: 10.1080/10749357.2015.1122266] [Medline: 27077973]
29. Thorsteinsen K, Vittersø J, Svendsen GB. Increasing physical activity efficiently: an experimental pilot study of a website and mobile phone intervention. Int J Telemed Appl 2014;2014:746232 [FREE Full text] [doi: 10.1155/2014/746232] [Medline: 24963290]
30. Zuckerman O, Gal-Oz A. Deconstructing gamification: evaluating the effectiveness of continuous measurement, virtual rewards, and social comparison for promoting physical activity. Pers Ubiquit Comput 2014 Jul 5;18(7):1705-1719. [doi: 10.1007/s00779-014-0783-2]
31. Direito A, Jiang Y, Whittaker R, Maddison R. Apps for IMproving FITness and increasing physical activity among young people: the AIMFIT pragmatic randomized controlled trial. J Med Internet Res 2015;17(8):e210 [FREE Full text] [doi: 10.2196/jmir.4568] [Medline: 26316499]
32. Edney SM, Olds TS, Ryan JC, Vandelanotte C, Plotnikoff RC, Curtis RG, et al. A social networking and gamified app to increase physical activity: cluster RCT. Am J Prev Med 2020 Feb;58(2):51-62. [doi: 10.1016/j.amepre.2019.09.009] [Medline: 31959326]
33. Garde A, Umedaly A, Abulnaga SM, Junker A, Chanoine JP, Johnson M, et al. Evaluation of a novel mobile exergame in a school-based environment. Cyberpsychol Behav Soc Netw 2016 Mar;19(3):186-192. [doi: 10.1089/cyber.2015.0281] [Medline: 26882222]
34. Gremaud AL, Carr LJ, Simmering JE, Evans NJ, Cremer JF, Segre AM, et al. Gamifying accelerometer use increases physical activity levels of sedentary office workers. J Am Heart Assoc 2018 Jul 02;7(13) [FREE Full text] [doi: 10.1161/JAHA.117.007735] [Medline: 29967221]

35. Höchsmann C, Müller O, Ambühl M, Klenk C, Königstein K, Infanger D, et al. Novel smartphone game improves physical activity behavior in type 2 diabetes. Am J Prev Med 2019 Jul;57(1):41-50 [FREE Full text] [doi: 10.1016/j.amepre.2019.02.017] [Medline: 31128953]
36. Kurtzman GW, Day SC, Small DS, Lynch M, Zhu J, Wang W, et al. Social incentives and gamification to promote weight loss: the LOSE IT randomized, controlled trial. J Gen Intern Med 2018 Oct;33(10):1669-1675. [doi: 10.1007/s11606-018-4552-1] [Medline: 30003481]
37. Helmerhorst HJ, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act 2012;9:103 [FREE Full text] [doi: 10.1186/1479-5868-9-103] [Medline: 22938557]
38. Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. Cochrane Handbook for Systematic Reviews of Interventions, Second Edition. Hoboken, New Jersey, United States: John Wiley & Sons; 2019.
39. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Med Res Methodol 2005 Apr 20;5:13 [FREE Full text] [doi: 10.1186/1471-2288-5-13] [Medline: 15840177]
40. DerSimonian R, Laird N. Meta-analysis in clinical trials revisited. Contemp Clin Trials 2015 Nov;45(Pt A):139-145 [FREE Full text] [doi: 10.1016/j.cct.2015.09.002] [Medline: 26343745]
41. IntHout J, Ioannidis JP, Borm GF. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerably outperforms the standard DerSimonian-Laird method. BMC Med Res Methodol 2014 Feb 18;14:25 [FREE Full text] [doi: 10.1186/1471-2288-14-25] [Medline: 24548571]
42. Hedges LV. Distribution theory for glass's estimator of effect size and related estimators. J Edu Stat 2016 Nov 23;6(2):107-128. [doi: 10.3102/10769986006002107]
43. Cohen J. Statistical Power for the Behavioral Sciences (2nd Edition). Hillsdale, NJ: Laurence Erlbaum Associates; 1988.
44. Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. Br Med J 2003 Sep 06;327(7414):557-560 [FREE Full text] [doi: 10.1136/bmj.327.7414.557] [Medline: 12958120]
45. Baujat B, Mahé C, Pignon J, Hill C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. Stat Med 2002 Sep 30;21(18):2641-2652. [doi: 10.1002/sim.1221] [Medline: 12228882]
46. Egger M, Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. Br Med J 1997 Sep 13;315(7109):629-634 [FREE Full text] [doi: 10.1136/bmj.315.7109.629] [Medline: 9310563]
47. Duval S, Tweedie R. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. Biometrics 2000 Jun;56(2):455-463. [doi: 10.1111/j.0006-341x.2000.00455.x] [Medline: 10877304]
48. Kossmeier M, Tran US, Voracek M. Power-enhanced funnel plots for meta-analysis. Zeitschrift für Psychologie 2020 Mar;228(1):43-49. [doi: 10.1027/2151-2604/a000392]
49. Harrer M, Cuijpers P, Furukawa TA, Ebert DD. Doing Meta-Analysis with R : A Hands-On Guide. Boca Raton, Florida, United States: Chapman and Hall/CRC; 2021.
50. McGuinness LA, Higgins JP. Risk-of-bias VISualization (robvis): an R package and Shiny web app for visualizing risk-of-bias assessments. Res Synth Methods 2021 Jan 06;12(1):55-61. [doi: 10.1002/jrsm.1411] [Medline: 32336025]
51. Garde A, Umedaly A, Abulnaga SM, Robertson L, Junker A, Chanoine JP, et al. Assessment of a mobile game ("MobileKids Monster Manor") to promote physical activity among children. Games Health J 2015 Apr;4(2):149-158. [doi: 10.1089/g4h.2014.0095] [Medline: 26181809]
52. Garde A, Chowdhury M, Rollinson AU, Johnson M, Prescod P, Chanoine JP, et al. A multi-week assessment of a mobile exergame intervention in an elementary school. Games Health J 2018 Feb;7(1):1-8. [doi: 10.1089/g4h.2017.0023] [Medline: 29394109]
53. Orwin R, Vevea J. Evaluating coding decisions. In: The Handbook of Research Synthesis and Meta-Analysis, 2nd Ed. New York: Russell Sage Foundation; 2009.
54. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000 Jan;55(1):68-78. [doi: 10.1037//0003-066x.55.1.68] [Medline: 11392867]
55. Prochaska JO, DiClemente CC. Transtheoretical therapy: toward a more integrative model of change. Psychotherapy: Theory, Research & Practice 1982;19(3):276-288. [doi: 10.1037/h0088437]
56. Bandura A. Health promotion from the perspective of social cognitive theory. Psychol Health 1998 Jul;13(4):623-649. [doi: 10.1080/08870449808407422]
57. Ajzen I. The theory of planned behavior. Organ Behav Hum Decis Process 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
58. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Appl Psychol 2008 Jan;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
59. Gomersall S, Maher C, English C, Rowlands A, Olds T. Time regained: when people stop a physical activity program, how does their time use change? A randomised controlled trial. PLoS One 2015 May;10(5):e0126665 [FREE Full text] [doi: 10.1371/journal.pone.0126665] [Medline: 26023914]
60. Lee I, Shiroma EJ, Kamada M, Bassett DR, Matthews CE, Buring JE. Association of step volume and intensity with all-cause mortality in older women. JAMA Intern Med 2019 May 29:1105-1112 [FREE Full text] [doi: 10.1001/jamainternmed.2019.0899] [Medline: 31141585]

61. Yates T, Zaccardi F, Dhalwani NN, Davies MJ, Bakrania K, Celis-Morales CA, et al. Association of walking pace and handgrip strength with all-cause, cardiovascular, and cancer mortality: a UK Biobank observational study. Eur Heart J 2017 Nov 14;38(43):3232-3240 [FREE Full text] [doi: 10.1093/eurheartj/ehx449] [Medline: 29020281]
62. Oja P, Kelly P, Murtagh EM, Murphy MH, Foster C, Titze S. Effects of frequency, intensity, duration and volume of walking interventions on CVD risk factors: a systematic review and meta-regression analysis of randomised controlled trials among inactive healthy adults. Br J Sports Med 2018 Jun;52(12):769-775. [doi: 10.1136/bjsports-2017-098558] [Medline: 29858464]
63. Murtagh EM, Nichols L, Mohammed MA, Holder R, Nevill AM, Murphy MH. The effect of walking on risk factors for cardiovascular disease: an updated systematic review and meta-analysis of randomised control trials. Prev Med 2015 Mar;72:34-43. [doi: 10.1016/j.ypmed.2014.12.041] [Medline: 25579505]
64. Gao H, Gao H, Sun FM, Zhang L. Effects of walking on body composition in perimenopausal and postmenopausal women: a systematic review and meta-analysis. Menopause 2016 Aug;23(8):928-934. [doi: 10.1097/GME.0000000000000627] [Medline: 27187009]
65. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. J Am Med Assoc 2018 Nov 20;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
66. Finkelstein EA, Haaland BA, Bilger M, Sahasranaman A, Sloan RA, Nang EE, et al. Effectiveness of activity trackers with and without incentives to increase physical activity (TRIPPA): a randomised controlled trial. Lancet Diabetes Endocrinol 2016 Dec;4(12):983-995. [doi: 10.1016/S2213-8587(16)30284-4] [Medline: 27717766]
67. Romeo A, Edney S, Plotnikoff R, Curtis R, Ryan J, Sanders I, et al. Can smartphone apps increase physical activity? Systematic review and meta-analysis. J Med Internet Res 2019 Mar 19;21(3):e12053 [FREE Full text] [doi: 10.2196/12053] [Medline: 30888321]
68. Chen XS, Changolkar S, Navathe AS, Linn KA, Reh G, Szwartz G, et al. Association between behavioral phenotypes and response to a physical activity intervention using gamification and social incentives: Secondary analysis of the STEP UP randomized clinical trial. PLoS One 2020;15(10):e0239288 [FREE Full text] [doi: 10.1371/journal.pone.0239288] [Medline: 33052906]
69. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
70. Faggiano F, Allara E, Giannotta F, Molinar R, Sumnall H, Wiers R, et al. Europe needs a central, transparent, and evidence-based approval process for behavioural prevention interventions. PLoS Med 2014 Oct;11(10):e1001740 [FREE Full text] [doi: 10.1371/journal.pmed.1001740] [Medline: 25291521]
71. Höchsmann C, Infanger D, Klenk C, Königstein K, Walz SP, Schmidt-Trucksäss A. Effectiveness of a behavior change technique-based smartphone game to improve intrinsic motivation and physical activity adherence in patients with type 2 diabetes: randomized controlled trial. JMIR Serious Games 2019 Feb 13;7(1):e11444 [FREE Full text] [doi: 10.2196/11444] [Medline: 30758293]
72. Haque MS, Kangas M, Jämsä T. A persuasive mhealth behavioral change intervention for promoting physical activity in the workplace: feasibility randomized controlled tria. JMIR Form Res 2020 May 04;4(5):e15083 [FREE Full text] [doi: 10.2196/15083] [Medline: 32364506]

Abbreviations

GRADE: grading of recommendations assessment, development, and evaluation
MD: mean difference
MET: metabolic equivalent of task
MVPA: moderate to vigorous physical activity
OSF: Open Science Framework
PA: physical activity
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial
SB: sedentary behavior
SMD: standardized mean difference

Edited by R Kukafka; submitted 25.12.20; peer-reviewed by C Maher, H Mehdizadeh, I Gupta, E Sadeghi-Demneh, H Pratomo, MDG Pimentel; comments to author 25.02.21; revised version received 31.03.21; accepted 19.10.21; published 04.01.22.

Please cite as:
Mazeas A, Duclos M, Pereira B, Chalabaev A
Evaluating the Effectiveness of Gamification on Physical Activity: Systematic Review and Meta-analysis of Randomized Controlled Trials
J Med Internet Res 2022;24(1):e26779
URL: https://www.jmir.org/2022/1/e26779
doi:10.2196/26779
PMID:34982715

Review

Facilitators and Barriers to the Adoption of Telemedicine During the First Year of COVID-19: Systematic Review

Clemens Kruse[1], MBA, MHA, MSIT, PhD; Katharine Heinemann[1], MHA

School of Health Administration, Texas State University, San Marcos, TX, United States

Corresponding Author:
Clemens Kruse, MBA, MHA, MSIT, PhD
School of Health Administration
Texas State University
Encino Hall, Room 250
601 University Dr
San Marcos, TX, 78666
United States
Phone: 1 5122454462
Fax: 1 5122458712
Email: scottkruse@txstate.edu

Abstract

Background: The virulent and unpredictable nature of COVID-19 combined with a change in reimbursement mechanisms both forced and enabled the rapid adoption of telemedicine around the world. Thus, it is important to now assess the effects of this rapid adoption and to determine whether the barriers to such adoption are the same today as they were under prepandemic conditions.

Objective: The objective of this systematic literature review was to examine the research literature published during the COVID-19 pandemic to identify facilitators, barriers, and associated medical outcomes as a result of adopting telemedicine, and to determine if changes have occurred in the industry during this time.

Methods: The systematic review was performed in accordance with the Kruse protocol and the results are reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. We analyzed 46 research articles from five continents published during the first year of the COVID-19 pandemic that were retrieved from searches in four research databases: PubMed (MEDLINE), CINAHL, Science Direct, and Web of Science.

Results: Reviewers identified 25 facilitator themes and observations, 12 barrier themes and observations, and 14 results (compared to a control group) themes and observations. Overall, 22% of the articles analyzed reported strong satisfaction or satisfaction (zero reported a decline in satisfaction), 27% reported an improvement in administrative or efficiency results (as compared with a control group), 14% reported no statistically significant difference from the control group, and 40% and 10% reported an improvement or no statistically significant difference in medical outcomes using the telemedicine modality over the control group, respectively.

Conclusions: The pandemic encouraged rapid adoption of telemedicine, which also encouraged practices to adopt the modality regardless of the challenges identified in previous research. Several barriers remain for health policymakers to address; however, health care administrators can feel confident in the modality as the evidence largely shows that it is safe, effective, and widely accepted.

(J Med Internet Res 2022;24(1):e31752) doi:10.2196/31752

KEYWORDS

telemedicine; pandemic; technology acceptance; COVID-19; digital health; telehealth; health policy; health care

Introduction

Rationale

The virulent nature of COVID-19 forced social distancing and a decrease of in-person visits to clinics around the world. Telemedicine presented health care providers with solutions that enabled a social-distancing window into the clinical environment and a continuation of the doctor-patient relationship.

Telemedicine is defined by the World Health Organization as healing from a distance through information communications technologies by all health care professionals for the "exchange of valid information for diagnosis, treatment and prevention of disease and injuries, research and evaluation" [1]. Telemedicine is not a perfect means of patient care; however, it offers great advantages to overcome geographical barriers to improve health outcomes [1]. Validated and peer-reviewed international statistics are elusive on adoption figures, but a recent question-and-answer session indicates overall low adoption of telemedicine internationally [2]. In the United States, prior to the pandemic, telemedicine had only been adopted by 8% of providers [3]. Providers have recognized wide acceptance of telemedicine by patients; however, prior to the desperate circumstances of COVID-19, they had not been willing to adopt telemedicine on a wide scale [4]. The largest challenges to the adoption of telemedicine were identified as technically challenged staff, resistance to change, cost, reimbursement, and education level of the patient [5]. Telemedicine saves patients time, consultation fees, and travel expenses [6]. However, telemedicine requires users at both ends to possess certain levels of technological skills such as those required to enable video teleconferencing [7]. Fortunately, some countries enacted legislation to expand the adoption of telemedicine. For example, in the United States, telemedicine was not easily reimbursed by federal programs until the Coronavirus Aid, Relief, and Economic Security (CARES) Act legislation [8], which greatly increased reimbursement mechanisms for the telemedicine modality. This change in reimbursement structure should not be ignored, and it most likely provided a significant catalyst to the adoption of telemedicine.

A large number of articles were published in the first 12 months of the pandemic (February 2020 to February 2021) on the rapid implementation efforts of telemedicine to enable clinics and hospitals to continue to see patients and care for their needs [9,10]. However, providers acknowledge some of the shortfalls inherent to this modality, such as lack of technical infrastructure, cost, lack of technical staff, computer literacy of both staff and patients, and a negative impact on the patient-to-provider relationship [4,11-13]. A systematic review performed in 2020 on telemedicine and COVID-19 evaluated 44 articles along four service lines and identified 10 themes of efficiency [14]. However, the authors did not evaluate facilitators and barriers to adoption or health outcomes. Another systematic review [5] was performed in 2016 on the barriers to the adoption of telemedicine worldwide, which evaluated 30 articles across all service lines in all countries; however, it also did not evaluate facilitators or health outcomes.

Although analyses have been published that highlight the advantages to the adoption of telemedicine, with an 8% adoption rate in the United States, the conclusions of these previous studies may not be as robust as possible. The circumstances presented by the pandemic have encouraged wider adoption of this modality of care. Therefore, with proper systematic review techniques, reviewer observations this far into the pandemic will undoubtedly be more robust and widely applicable to medicine.

Objectives

The purpose of this systematic review was to evaluate the facilitators and barriers to the adoption of telemedicine worldwide, including an analysis of health outcomes and patient satisfaction. A brief comparison of the results of this review with those of reviews performed prior to COVID-19 was further performed to identify changes in these factors in light of the pandemic.

Methods

Protocol and Registration

The Kruse protocol for writing a systematic review was followed, and the findings are reported in accordance with the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines [15,16]. This systematic review was registered in PROSPERO on August 2, 2021 (ID CRD42021235933).

Eligibility Criteria

The search parameters were established to find articles published in 2020 and 2021 concerning telemedicine in all aspects of care and for all ages of patients, published in peer-reviewed journals, using any method of study (mixed method, quantitative, and qualitative). Other systematic reviews were excluded because we wanted to compare our results to these previous reviews without confounding the findings. The Johns Hopkins Nursing Evidence-Based Practice Rating Scale (JHNEBP) was used to assess the quality of all articles analyzed [17]. Any studies below level IV C were discarded due to poor quality.

Information Sources

Four research databases were searched: PubMed (MEDLINE), CINAHL (excluding MEDLINE), Web of Science, and Science Direct. We also performed a journal-specific search of the Journal of Medical Internet Research.

Search Strategy

Google Scholar was used to determine the general trends of publication on this topic previously and to collect key terms from published articles. These key terms were entered into the US Library of Medicine's Medical Subject Headings (MeSH) to create an exhaustive search string using Boolean terms. The actual search string used was: (telemedicine OR telehealth OR "mobile health" OR mhealth OR ehealth) AND (COVID-19 OR coronavirus). The same search string was used in all databases. Similar filters were used in each database (not all filters are the same between databases).

Study Selection Process

Once the search string was entered into each database, we filtered the results and screened abstracts for applicability. Although filters for the four research databases differ, we generally filtered for the date range (2020-2021), scholarly journals (no theses or opinions), and "full text" to ensure that we would have access to the entire article. Articles were rejected for a variety of reasons: protocol (no results to analyze); opinion (no data); reviews; did not use telemedicine; or did not contribute to our objective statement of identifying facilitators, barriers, or effects on patient satisfaction. The κ statistic was calculated to identify the level of agreement between reviewers [18].

Data Collection Process

An Excel spreadsheet was used as a data-extraction tool to collect data for reporting and analysis. This spreadsheet was standardized according to the Kruse protocol [15]. We held three consensus meetings to screen abstracts, analyze articles, and discuss possible themes. After the second consensus meeting, we performed a narrative analysis to identify themes in the articles analyzed [19]. Because there were only two authors on this project, both authors analyzed all articles (n=46).

Data Items

In accordance with the Kruse protocol, PRISMA standard, and JHNEBP, the following fields were collected: database source; date of publication; journal; authors; study title; PICOS (participants, intervention, results, outcomes, study design); sample size; bias within study; effect size; country of origin; statistics used; quality metrics from the JHNEBP scale; and reviewer observations as they relate specifically to the objective statement in areas of patient satisfaction, and facilitators and barriers to adoption [15,17,20]. All data items were independently collected and discussed in subsequent consensus meetings.

Risk of Bias Within and Across Studies

The JHNEBP rating scale was used for assessment of bias within and across studies. Observations of bias and methodological weaknesses were noted [17]. The JHNEBP ratings also provided insight into bias because poor-quality results can limit the external validity of the experiment.

Summary Measures

Because we included mixed methods and qualitative studies, we were unable to standardize summary measures as would be performed in a meta-analysis.

Additional Analyses

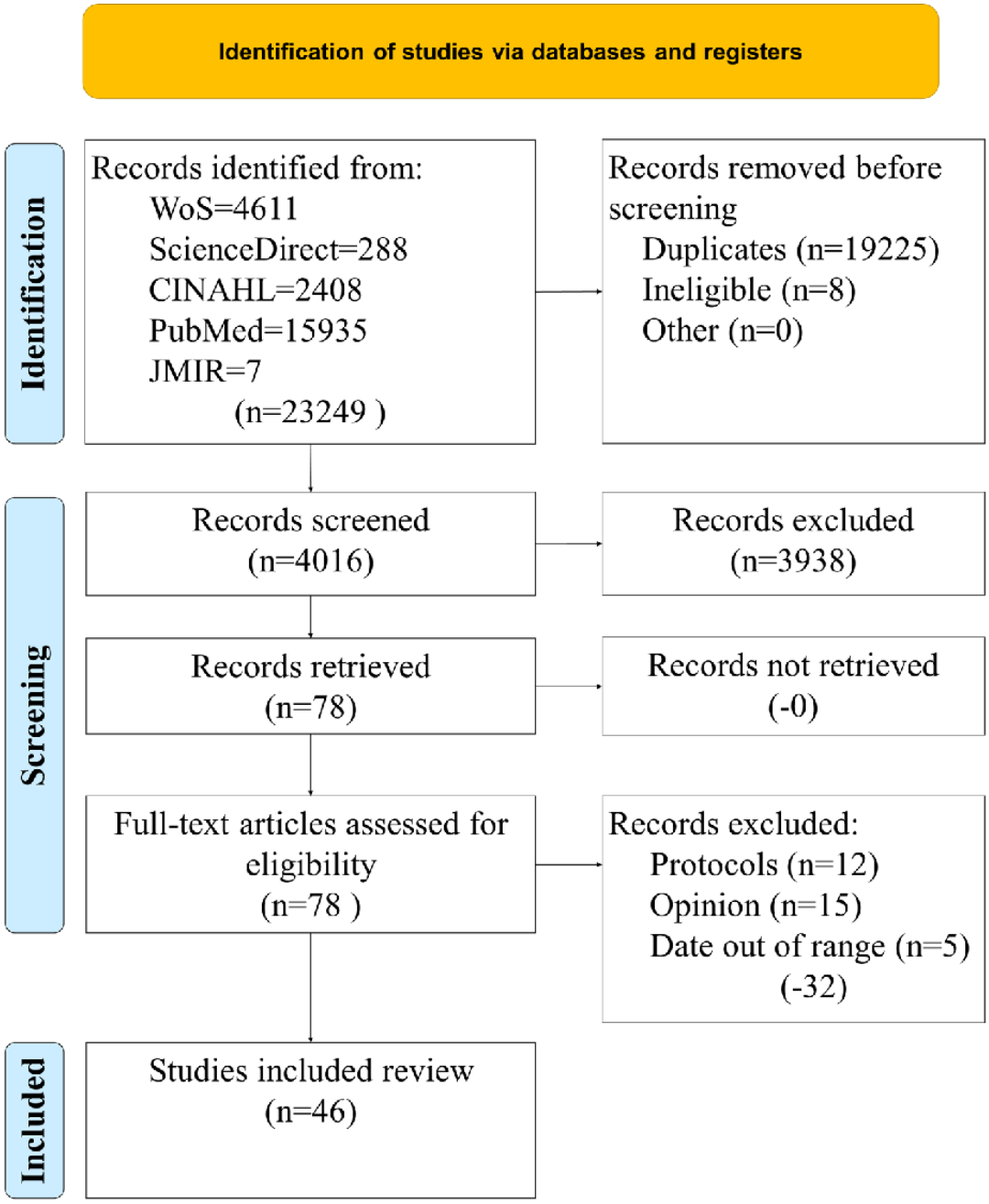
We performed a narrative, or thematic, analysis of the observations to convert them into themes (common threads between articles) [19]. We calculated the frequency of occurrence of both themes and individual observations and report these in a series of affinity matrices (tables). This technique was used to identify the statistical probability for identifying each theme, which does not identify a level of importance but rather identifies a frequency of mention of these themes in the literature during the period of observation.

Results

Study Selection

The database search and study selection process are illustrated in Figure 1. The κ statistic was 0.95, indicating almost perfect agreement between reviewers [18,21]. Several studies made it through all filters, but were still eliminated because they were protocols (no results), opinions, out of the date range, or other systematic reviews.

Figure 1. Article search and selection process. WoS: Web of Science.



Study Characteristics

Reviewers collected study characteristics identified by the PRISMA standard such as PICOS (see Table 1). Of the 46 six studies analyzed over the 15-month period, 2 (4%) involved adolescents, 6 (13%) involved adults >60 years, and 38 (83%) involved adults >18 years as participants. Most participants were current or former patients who agreed to participate in studies. More than half the interventions were mobile health (mHealth), telephone/televideo, or eHealth (26/46, 56%). The rest were interventions involving telemonitoring, patient portals, telecoaching, web chat, and social media, which could be cross-platform. In these 46 studies, 18 resulted in a positive outcome over a control group (23%), 12 of which involved medically measured outcomes (21%) as opposed to clinical and administrative outcomes. Only 9 of the 46 (20%) studies resulted in no statistically significant difference between the intervention and control groups, which means that positive results could be obtained through telemedicine commensurate with those obtained using traditional means of care. Four articles analyzed were published in 2021 [22-25], with the remaining 42 articles published in 2020 [26-67]. Further explanation of the results and medical outcomes can be found the Additional Analysis subsection.

Table 1. Characteristics of the included studies according to the PICOS (Participants, Intervention, Results, Outcomes, and Study Design) structure.

Study	Participant	Intervention	Results (compared to the control group or other studies)	Medical outcomes	Design
Ben-Arye et al [22]	Adult patients (>18 years) undergoing adjuvant, neoadjuvant, or palliative treatment for solid tumors	eHealth	Improved compliance/adherence	Not reported	Prospective, controlled, and nonrandomized study
Yu et al [25]	Older adult patients (50% >60 years, 60% women, 68% one-time telehealth users) and 45 physicians	Telephone or televideo	Improved patient satisfaction	Not reported	Cross-sectional
Richards et al [24]	Adult respondents from a neurosurgical outpatient clinic (mean age 63 years, 50.3% men)	Telephone or televideo	Improved patient satisfaction	Not reported	Qualitative
Kurihara et al [23]	Adult patients with Parkinson disease (61% women, mean age 67 years) at Fukuoka University Hospital	Telemedicine self-testing	No control group (nonexperimental)	Not reported	Cross-sectional
Alkire et al [26]	Adults (Gen X, Millennial)	Patient portals	No control group (nonexperimental)	Not reported	Nonexperimental
Ballin et al [27]	Older adults, 70-year-old men, and women with central obesity	Supervised and web-based	No significant difference; decreased fat mass	Improved in at least one area: decreased fat mass	Randomized controlled trial
Banbury et al [28]	Adults >50 years with at least one chronic condition	Telemonitoring	Telemedicine improved results compared to control: companionship, emotional support, health literacy, self-management	Not reported	Mixed methods, quasi-experimental, nonrandomized trial
Barnett et al [29]	Adults (22-27 years; 10 men, 10 women), clients of an alcohol and drug counseling service across Australia, and 8 counselors	Webchat	No control group (nonexperimental)	Not reported	Qualitative study, nonexperimental
Batalik et al [30]	Adult cardiac rehabilitation patients	Home-based telerehab	No statistically significant difference	No statistically significant difference	Randomized controlled trial
Beller et al [31]	Adult patients scheduled for video visits through the University of Virginal urology departments	Televideo	No control group (nonexperimental)	Not reported	Cohort
Bernabe-Ortiz et al [32]	Adult participants from a randomized clinical trial on a 1-year mHealth[a] intervention on blood pressure and body weight 4 years postcompletion	mHealth	Telemedicine improved results compared to control: decreased fat mass	Improved in at least one area; decreased body weight	Retrospective study of a randomized clinical trial
Bilgrami et al [33]	Adults with inflammatory bowel disease	Telemedicine self-testing	No statistically significant difference	No statistically significant difference	Randomized controlled trial
Broers et al [34]	Adult patients with cardiovascular disease	eHealth	No statistically significant difference; increased quality of life	Not reported	Randomized controlled trial

Study	Participant	Intervention	Results (compared to the control group or other studies)	Medical outcomes	Design
Cho et al [35]	Adult participants (30-59 years) with at least 2 conditions defined by the Third Report of the National Cholesterol Education Program expert panel (abdominal obesity, high blood pressure, high triglycerides, low high-density lipoprotein cholesterol, and high fasting glucose level)	mHealth	Telemedicine improved results compared to control (decreased fat mass)	Improved in at least one area: decreased fat mass, decreased body weight	Randomized controlled trial
Claes et al [36]	Adult patients with cardiovascular disease from 3 European hospitals	eHealth	Improved health behaviors	Not reported	Randomized controlled trial
Coorey et al [37]	Adults who had completed 12 months of follow-up from the Consumer Navigation of Electronic Cardiovascular Tools trial	eHealth	No control group (nonexperimental): improved self-management, improved health literacy	Not reported	Qualitative analysis of a randomized controlled study
Ding et al [38]	Adults (mean age 70.1 years) with chronic heart failure	Telemonitoring	Telemedicine improved results compared to controls: improved compliance/adherence	Not reported	Randomized controlled trial
Geramita et al [39]	Adult lung transplant recipients	mHealth	No statistically significant difference	Not reported	Randomized controlled follow-up study
Gong et al [40]	Adult hypertension	mHealth	Telemedicine improved results compared to controls: improved compliance/adherence	Improved in at least one area: reductions in blood pressure	Randomized controlled trial
Han et al [41]	Adults (<55 years) prepandemic (S1) and 273 follow-up surveys (S2); university-affiliated, and physicians	eHealth	No control group (nonexperimental): telemedicine improved results compared to controls, improved compliance/adherence	Not reported	Qualitative
Harding et al [42]	Adult caregivers with 837 patient assessment outcomes	mHealth	No control group (nonexperimental)	Not reported	Qualitative (pilot study)
Hsia et al [43]	Pediatric patients with asthma	mHealth	Telemedicine improved results compared to controls: improved self-management, improved patient satisfaction	Improved self-management, decreased medication use, increase in controlled asthma	Prospective study
Hsieh et al [44]	Insured adults (>20 years)	Patient portals	No control group (nonexperimental)	Not reported	Qualitative
Hutchesson et al [45]	Adult Australian women with a recent history of preeclampsia	mHealth	No statistically significant difference	No statistically significant difference	Pilot randomized controlled trial
Jiménez-Marrero et al [46]	Adult patients with chronic heart failure	Televideo	Telemedicine improved results compared to controls, decreased cost	Improved in at least one area: decreased incidence of heart failure	Randomized controlled trial

Study	Participant	Intervention	Results (compared to the control group or other studies)	Medical outcomes	Design
Katt et al [47]	180 patients with upper-extremity condition and 302 physicians	Telephone or televideo	Improved patient satisfaction	Not reported	Qualitative
Kobe et al [48]	Adult patients (52% men, mean age 62 years, 55.5% African American) of Duke University Health System with type 2 diabetes, poorly controlled hypertension, and on prescription hypertension and diabetes medication	Telephone or televideo	Telemedicine improved results compared to control	Improved in at least one area, improved annual rate eGFR[b] decline	Secondary analysis of randomized controlled trial
Lai et al [49]	Adults with Parkinson disease (telehealth mean age 63 years, control mean age 70 years; 70% men, predominantly White)	Telemonitoring	Telemedicine improved results compared to control: improved compliance/adherence, health behaviors, and patient satisfaction	Not reported	Mixed methods
Lemelin et al [50]	Adult women (mean age 32 years) with gestational diabetes mellitus	Telecoaching	Improved patient satisfaction: telemedicine improved results compared to control	Identified other areas for intervention	Prospective and controlled clinical trial
Manning et al [51]	Adults from families with toddlers	Televideo	No statistically significant difference	Not reported	Mixed method quasiexperimental and longitudinal design
Marques et al [52]	Adult Valladolid University students (74% women, 67.5% aged 18-23 years)	mHealth	No control group (nonexperimental)	Not reported	Qualitative
Martins et al [53]	Adult patients (mean age 62 years, 50% women) with suspected acute strokes at a Brazil university hospital	mHealth	Telemedicine improved results compared to control	Improved in at least one area: decreased mortality, decreased intracranial hemorrhage	Prospective observational
McGillicuddy et al [54]	Adults (mean 51.5-52.1 years) with kidney transplants (majority men, African American)	mHealth	Telemedicine improved results compared to control	Improved in at least one area: reduction in mean tacrolimus trough coefficient of variation	Randomized controlled clinical trial
Mo et al [55]	Adult patients (51.7-53.5 years) with chronic heart failure (approximately 66% men)	Telephone or televideo	Telemedicine improved results compared to control: improved emotional support	Improved in at least one area: mental health inventory, quality of life	Open-label interventional study
Mustonen et al [56]	Adult patients (>45 years; mean age 65 years) with type 2 diabetes and coronary artery disease (approximately 40% women)	Telecoaching	No statistically significant difference	Not reported	Posttrial analysis of a randomized controlled trial
O'Shea et al [57]	Adults (77% men, mean age 61 years)	eHealth	Not reported	Not reported	Posttrial analysis of an acceptability and feasibility trial
Perri et al [58]	Adults (mean 55.4 years) from 14 counties in Florida (83% women, 73.9% White)	Telephone or televideo	Telemedicine improved results compared to control: decreased fat mass, improved self-management	Improved in at least one area: decreased body weight	Randomized clinical trial

Study	Participant	Intervention	Results (compared to the control group or other studies)	Medical outcomes	Design
Piera-Jiménez et al [59]	Adults (majority 50-70 years and men) from Spain, the Netherlands, and Taiwan	Telemonitoring	Telemedicine improved results compared to control	Improved in at least one area, improved quality of life	Financial randomized controlled trial
Press et al [60]	Adults (mean 54.5 years) with asthma or chronic obstructive pulmonary disease (majority Black women)	mHealth	Telemedicine improved results compared to control: improved self-management health behaviors	Increase in controlled asthma	Randomized controlled trial
Ramirez-Correa et al [61]	Adults (mean 39.9 years, 56% men)	Telemedicine self-testing	No control group (nonexperimental)	Not reported	Cross-sectional
Ronan et al [62]	Adults with cystic fibrosis involved in a study on an online Tai Chi intervention	Televideo	No statistically significant difference, improved health behaviors	Not reported	Qualitative analysis of a mixed methods randomized controlled feasibility study
Sacco et al [63]	Older adults (mean age 88.2 years), 59.8% women	Telephone or video	Improved patient satisfaction, improved emotion support	Not reported	Cross-sectional survey
Scheerman et al [64]	Adolescents (12-17 years) and mothers	Social media	Telemedicine improved results compared to control, improved health behaviors	Not reported	Cluster randomized controlled trial
Schrauben et al [65]	Adult Chronic Renal Insufficiency Cohort (CRIC) Study participants (mean age 68 years, eGFR 54 mL/min/1.73, 59% men)	mHealth	No control group (nonexperimental)	Not reported	Cross-sectional survey
Shareef et al [66]	Elderly and disabled people (average age 74.5 years, 59% women) in retirement homes and rehabilitation centers	Robotics or artificial intelligence	Improved companionship	Not reported	Experiment and follow-up survey
van Dijk et al [67]	Adult women (mean age 30 years), either less than 13 weeks pregnant or trying to become pregnant, and 36 men	mHealth	Improved compliance/adherence, improved health behaviors	Improved in at least one area, improved self-management	Randomized controlled trial

[a]mHealth: mobile health.

[b]eGFR: estimated glomerular filtration rate.

Risk of Bias Within and Across Studies

Table 2 summarizes the quality indicators assessed for each article with the JHNEBP tool. The strength of evidence most frequently observed was level III followed by level I and level II. Nearly half of the articles reported strong-evidence studies that included both a control group and randomization; the next most common study type was nonexperimental (no control group) or qualitative, with the least frequent type being quasiexperimental (included a control group but no randomization). The quality of evidence most frequently observed was A (high quality), followed by B (good quality). The most common combination of strength and quality was III B, followed closely by I A, which speaks to both the strength and quality of evidence evaluated by this review. The III B combination highlights the number of qualitative studies with smaller samples or selection bias.

Table 2. Summary of quality assessments (N=46).

Evidence	Occurrence, n (%)
Strength	
I (Experimental study or randomized controlled trial)	22 (48)
III (Nonexperimental, qualitative)	17 (37)
II (quasiexperimental)	7 (15)
Quality	
A (High quality)	27 (59)
B (Good quality)	17 (37)
C (Low quality)	2 (4)

Many studies used geographically localized samples, which may limit the external validity of the results. Some studies focused only on one gender or race, speaking to the convenience sample or volunteer-basis of their design. Asking for volunteers in a technology-oriented experiment invites bias because the self-selection allows for those who are already technology-oriented or comfortable with technology to participate. This group as the intervention can skew the results because those already comfortable with technology will not experience the frustration experienced by those who are not comfortable with technology. This selection bias also limits the external validity of the results. A comprehensive list of bias, country of origin, sample size, strength, and quality of evidence identified for each study can be found in Multimedia Appendix 1.

Thematic Analysis Based on Results of Individual Studies

During the analysis phase of the systematic review process, the reviewers recorded observations to identify instances of patient satisfaction, as well as both facilitators and barriers to the adoption of telemedicine. A thematic analysis was then performed to make sense of the observations [19]. Multiple instances of the same observation become a theme. A translation of observations to themes is provided in Multimedia Appendix 2. The summary of analysis is provided in Table 3, which lists the themes/observations from reviewers that correspond with the objective statement and sorts articles from the most recent to the oldest.

Table 3. Summary of thematic analysis for individual studies.

Authors	Patient satisfaction	Facilitators	Barriers
Ben-Arye et al [22]	Not reported	Technical literacy, availability of technology, past experience with technology	Availability of technology, confidentiality/security
Yu et al [25]	Strong satisfaction	Concerns adequately addressed, improved health behaviors, pandemic created acceptance of technology	Some patients prefer in-person consultations, decrease in patient-provider communication, technical literacy
Richards et al [24]	Strong satisfaction	Convenience of telemedicine, increased patient-provider communication, concerns adequately addressed, increased access	Not reported
Kurihara et al [23]	Not reported	Pandemic created acceptance of technology, past experience with technology	Some patients prefer in-person consultations, technical literacy
Alkirie et al [26]	Not reported	Technical literacy, past experience with technology, perceived usefulness, increased patient-provider communication, perceived ease of use	Technology needs further development, technical literacy
Ballin et al [27]	Not reported	Increased connectedness, self-management, flexibility, and access	Technology needs further development
Banbury et al [28]	Not reported	Enabled social interaction; decreased anxiety; increased connectedness, technical literacy, and access; televideo enables reading of body language; education; convenience of telemedicine	Health literacy, availability of technology, technical literacy
Barnett et al [29]	Not reported	Increased efficiency, access, and patient-provider communication, and improved standard of care	Technology needs further development, decrease in patient-provider communication, technical literacy, confidentiality/security
Batalik et al [30]	Not reported	Technical literacy, increased self-management, increased access, increased flexibility	Discomfort for wearable monitors, technical literacy, technology needs further development
Beller et al [31]	Not reported	Pandemic created acceptance of technology, availability of technology, fewer miles driven to appointment, convenience of telemedicine, faster initiation of treatment, decreased costs	Limits of reimbursement for telemedicine, some patients prefer in-person consultations, connectivity, technical literacy
Bernabe-Ortiz et al [32]	Not reported	Increased connectedness, increased adherence, improved health behaviors	Perceived lack of usefulness, lack of personal desire to get better, some patients prefer in-person consultations
Bilgrami et al [33]	Not reported	Pandemic created acceptance of technology	Not reported
Broers et al [34]	Strong satisfaction	Perceived usefulness, perceived ease of use, increased adherence	Decrease in quality of life after intervention
Cho et al [35]	Not reported	Increased adherence, increased self-management, increased weight loss, technical literacy	Technical literacy, availability of technology
Claes et al [36]	Not reported	Technical literacy, perceived ease of use	Technology needs further development
Coorey et al [37]	Not reported	Increased adherence, increased self-management	Lack of personal desire to get better, technology needs further development, technical literacy
Ding et al [38]	Not reported	Increased adherence, increased self-management	Technology needs further development, cost
Geramita et al [39]	Not reported	Long-term use may not be required to develop good habits	Cost, confidentiality/security, technology needs further development
Gong et al [40]	Not reported	Increased adherence, increased self-management	Not reported
Han et al [41]	Not reported	Pandemic created acceptance of technology, increased efficiency, increased self-management, increased access, availability of technology	Cost, technical literacy, interoperability, availability of technology
Harding et al [42]	Not reported	Not reported	Connectivity, confidentiality/security, technical literacy
Hsia et al [43]	Strong satisfaction	Increased quality of life, decreased emergency room visits, increased adherence, availability of technology, pandemic created acceptance of technology, perceived ease of use, convenience of telemedicine	Connectivity, technical literacy, cost, availability of technology

Authors	Patient satisfaction	Facilitators	Barriers
Hsieh et al [44]	Not reported	Health literacy, perceived usefulness, perceived ease of use	Some patients prefer in-person consultations, technical literacy, cost
Hutchesson et al [45]	Strong satisfaction	Increased self-management, perceived usefulness, perceived ease of use	Technology needs further development, perceived lack of usefulness
Jiménez-Marrero et al [46]	Not reported	Decreased costs, increased adherence, increased self-management	Cost
Katt et al [47]	Strong satisfaction	Convenience of telemedicine, pandemic created acceptance of technology, faster initiation of treatment, perceived ease of use	Some patients prefer in-person consultations, workflow issues for providers
Kobe et al [48]	Not reported	Not reported	Some patients prefer in-person consultations
Lai et al [49]	Strong satisfaction	Convenience of telemedicine, increased social support, increased self-management	Technology needs further development, connectivity, decrease in patient-provider communication, technical literacy
Lemelin et al [50]	Strong satisfaction	Education, increased social support	Not reported
Manning et al [51]	Not reported	Pandemic created acceptance of technology	Connectivity, availability of technology
Marquez et al [52]	Not reported	Past experience with technology, decreased costs, pandemic created acceptance of technology, faster initiation of treatment, increased access	Some patients prefer in-person consultations
Martins et al [53]	Not reported	Faster initiation of treatment, availability of technology, increased access	Lack of infrastructure, limits of reimbursement for telemedicine, connectivity, confidentiality/security
McGillicuddy et al [54]	Not reported	Increased social support, health literacy	Not reported
Mo et al [55]	Not reported	Increased quality of life, increased social support	Not reported
Mustonen et al [56]	Not reported	Decreased costs	Not reported
O'Shea et al [57]	Satisfaction	Increased self-management	Technical literacy, perceived lack of usefulness, technology needs further development
Perri et al [58]	Not reported	Increased weight loss, increased adherence, increased self-management	Not reported
Piera-Jiménez et al [59]	Not reported	Decreased costs, no significant difference in cost care	Cost
Press et al [60]	Not reported	Decreased costs, education, increased access	Availability of technology, technical literacy
Ramirez-Correa et al [61]	Not reported	Increased patient-provider communication, education, pandemic created acceptance of technology	Connectivity
Ronan et al [62]	Not reported	Convenience of telemedicine, pandemic created acceptance of technology, increased social support	Technical literacy, technology needs further development, availability of technology
Sacco et al [63]	Strong satisfaction	Increased social support, increased connectedness	Not reported
Scheerman et al [64]	Not reported	Increased social support, improved standard of care	Not reported
Schrauben et al [65]	Not reported	Health literacy, education	Technical literacy, health literacy, confidentiality/security
Shareef et al [66]	Not reported	Enabled social interaction, increased social support	Confidentiality/security, technical literacy, perceived lack of usefulness
van Dijk et al [67]	Not reported	Improved health behaviors, increased adherence	Not reported

Patient satisfaction was reported as "strong satisfaction" or "satisfaction" in 9 (20%) and 1 (2%) of the 46 studies, respectively, and 36 studies did not report any measure of patient satisfaction. No studies reported a decline in patient satisfaction as a result of using telemedicine as the intervention.

Twenty-five facilitator themes and seven individual observations were identified in the literature by the two reviewers. Only two studies did not identify facilitators. Facilitator themes are listed in Table 4.

Table 4. Facilitator themes and individual observations (N=132).

Themes/observations	References	Occurrence, n (%)
Increased self-management	[27,30,35,37,38,40,41,45,46,49,57,58]	12 (9.1)
Pandemic created acceptance of technology	[23,25,31,33,41,43,47,51,52,61,62]	11 (8.3)
Increased adherence	[32,34,35,37,38,40,43,46,58,67]	10 (7.6)
Increased access	[24,27-30,41,52,53,60]	9 (6.8)
Increased social support	[49,50,54,55,62-64,66]	8 (6.1)
Convenience of telemedicine	[24,28,31,43,47,49,62]	7 (5.3)
Perceived ease of use	[26,34,36,43-45,47]	7 (5.3)
Decreased costs	[31,46,52,56,59,60]	6 (4.5)
Education	[28,50,60,61,65]	5 (3.8)
Technical literacy	[22,26,30,35,36]	5 (3.8)
Availability of technology	[22,31,41,43,53]	5 (3.8)
Increased patient-provider communication	[24,26,29,61]	4 (3.0)
Faster initiation of treatment	[31,47,52,53]	4 (3.0)
Increased connectedness	[27,28,32,63]	4 (3.0)
Perceived usefulness	[26,34,44,45]	4 (3.0)
Past experience with technology	[22,23,26,52]	4 (3.0)
Health literacy	[44,54,65]	3 (2.3)
Improved health behaviors	[25,32,67]	3 (2.3)
Increased efficiency	[29,41]	2 (1.5)
Concerns adequately addressed	[24,25]	2 (1.5)
Enabled social interaction	[28,66]	2 (1.5)
Increased quality of life	[43,55]	2 (1.5)
Improved standard of care	[29,64]	2 (1.5)
Increased flexibility	[27,30]	2 (1.5)
Increased weight loss	[35,58]	2 (1.5)
Decreased anxiety	[28]	1 (0.8)
Increased technical literacy	[28]	1 (0.8)
Televideo enables reading of body language	[28]	1 (0.8)
Fewer miles driven to appointment	[31]	1 (0.8)
Long-term use may not be required to develop good habits	[39]	1 (0.8)
Decreased emergency room visits	[43]	1 (0.8)
No significant difference in cost of care	[59]	1 (0.8)
Not reported	[42,48]	2 (N/A[a])

[a]N/A: not applicable.

The most commonly identified themes were increased self-management, acceptance of the technology from the pandemic, adherence to treatment protocols, access, and social support. For the 46 articles, these themes represent 38% of all 132 occurrences. Other themes included convenience of telemedicine and perceived ease of use, decreased cost, opportunity for education, technical literacy, availability of technology, an increase in patient-provider communication, faster initiation of treatment, increased connectedness, perceived usefulness, and past experience with technology. Health literacy and improved health behaviors were identified less frequently, and increased office efficiencies, medical concerns adequately addressed, enabled social interaction, increased quality of life, improved standard of care, increased flexibility, and increased weight loss were the least frequent themes identified. The following seven individual observations accounted for 5% of the total observations: decreased anxiety, increased technical literacy, televideo enabled reading of body language, fewer miles driven to appointment, long-term use may not be required

to develop good habits, decreased emergency room visits, and no significant difference in cost of care.

Twelve themes and five individual observations were identified as barriers from the literature by the reviewers; 11 studies did not identify barriers (11%). Table 5 lists the themes and individual observations.

The most commonly listed barriers were technical literacy, technology needs further development, availability of technology, and patient preference, accounting for 55% of the total 86 occurrences. Cost, connectivity, and confidentiality/security were also identified, as well as health literacy, limits of reimbursement for telemedicine, and lack of personal desire to get better with less frequent occurrences (2 each). The remaining five observations made up a total of 6% of the total occurrences: decrease in quality of life after intervention, discomfort for wearing monitors, workflow issues for providers, lack of data infrastructure, and interoperability.

Table 5. Barrier themes and individual observations (N=86).

Themes/observations	References	Occurrence, n (%)
Technical literacy	[23,25,26,28-31,35,37,41-44,49,57,60,62,65,66]	19 (22)
Technology needs further development	[26,27,29,30,36-39,45,49,57,62]	12 (14)
Availability of technology	[22,28,35,41,43,51,60,62]	8 (9)
Cost	[38,39,41,43,44,46,59]	7 (8)
Connectivity	[31,42,43,49,51,53,61]	7 (8)
Confidentiality/security	[22,29,39,42,53,65,66]	7 (8)
Some patients prefer in-person consultations	[23,25,31,32,44,47,48,52]	8 (9)
Perceived lack of usefulness	[32,45,57,66]	4 (5)
Decrease in patient-provider communication	[25,29,49]	3 (3)
Health literacy	[28,65]	2 (2)
Limits of reimbursement for telemedicine	[31,53]	2 (2)
Lack of personal desire to get better	[32,37]	2 (2)
Decrease in quality of life after intervention	[34]	1 (1)
Discomfort for wearable monitors	[30]	1 (1)
Workflow issues for providers	[47]	1 (1)
Lack of infrastructure	[53]	1 (1)
Interoperability	[41]	1 (1)
Not reported	[24,33,40,50,54-56,58,63,64,67]	11 (N/A[a])

[a]N/A: not applicable.

Additional Analyses

Distribution of Publications by Country

Eighteen of the 46 studies (39%) were performed in North America, 11 (24%) were performed in Europe, 7 (15%) were performed in Asia, 5 (11%) were performed in Australia, 3 (7%) were performed in South America, and 2 (4%) were performed in multiple countries and continents.

Comparisons to a Control Group

Table 6 summarizes the themes and observations recorded for results as compared to the control group identified by the two reviewers. There is some overlap between this set of observations and medical outcomes; the latter represent clinical observations only, whereas the former are both clinical and administrative in nature. Ten themes and four individual observations were identified by the reviewers for a total of 66 occurrences in the literature. Eleven studies were nonexperimental in nature, which had no control group.

Eighteen of the studies demonstrated either a clinical or administrative improvement compared to the control group, whereas nine reported no statistically significant results from the control group. Both of these themes demonstrate the efficacy of the telemedicine modality. The remainder of the list in Table 6 demonstrates the specific improvements that occurred (multiple improvements occurred in multiple articles), including improved patient satisfaction, improved behaviors, improved compliance/adherence to treatment protocol, improved self-management of condition or disease, decreased fat mass, improved emotional support, improved companionship, and improved health literacy. The remainder were individual observations that combined accounted for 5% of the total observations: improved informational support, decreased cost, and increased quality of life. Only one article did not report a result as compared to the control group because it was a posttrial analysis and it did not address the control group.

Table 6. Themes and individual observations for studies with a control group comparison (N=66).

Themes/observations	References	Occurrence, n (%)
Telemedicine improved results compared to control	[28,32,35,38,40,41,43,46,48-50,53-55,58-60,64]	18 (27)
No statistically significant difference	[27,30,33,34,39,45,51,56,62]	9 (14)
Improved patient satisfaction	[24,25,43,47,49,50,63]	7 (11)
Improved health behaviors	[36,49,60,62,64,67]	6 (9)
Improved compliance/adherence	[22,38,40,41,49,67]	6 (9)
Improved self-management	[28,37,43,58,60]	5 (8)
Decreased fat mass	[27,32,35,58]	4 (6)
Improved emotional support	[28,55,63]	3 (5)
Improved companionship	[28,66]	2 (3)
Improved health literacy	[28,37]	2 (3)
Improved informational support	[28]	1 (2)
Decreased cost	[46]	1 (2)
Increased quality of life	[34]	1 (2)
Not reported	[57]	1 (2)
No control group (nonexperimental)	[23,26,29,31,37,41,42,44,52,61,65]	11 (N/A[a])

[a]N/A: not applicable.

Medical Outcomes Commensurate With an Intervention

Table 7 summarizes the medical outcomes observed. Seven themes and nine individual observations were recorded commensurate with the adoption of telemedicine for a total of 30 occurrences. Twenty-eight studies did not report clinical outcomes.

Twelve studies reported 12 statistically significant improvements in clinical outcomes and three reported no statistically significant difference between modalities of care. Both of these themes demonstrated the efficacy of the telemedicine modality. The most commonly observed theme for medical outcomes was decreased body weight, followed by decreased fat mass, improved self-management, increase in controlled asthma, and increased quality of life. The following individual observations contributed to 30% of the total observations: reduction in blood pressure, reduction in mean tacrolimus trough coefficient of variation, improved annual rate of estimated glomerular filtration rate (eGFR) decline, decrease in medication use, decrease incidence of heart failure, decreased mortality, improved mental health inventory, decreased intracranial hemorrhage, and telemedicine identified other areas for intervention.

Table 7. Medical outcome themes and individual observations commensurate with adoption of the intervention/technology (N=30).

Themes/observations	References	Occurrence, n (%)
Improved in at least one area	[27,32,35,40,46,48,53-55,58,59,67]	12 (40)
No statistically significant difference	[30,33,45]	3 (10)
Decreased body weight	[32,35,58]	3 (10)
Decreased fat mass	[27,35]	2 (7)
Improved self-management	[43,67]	2 (7)
Increase in controlled asthma	[43,60]	2 (7)
Improved quality of life	[55,59]	2 (7)
Reductions in blood pressure	[40]	1 (3)
Reduction in mean tacrolimus trough coefficient of variation	[54]	1 (3)
Improved annual rate of eGFR[a] decline	[48]	1 (3)
Decreased medication use	[43]	1 (3)
Decreased incidence of heart failure	[46]	1 (3)
Identified other areas for intervention	[50]	1 (3)
Decreased mortality	[53]	1 (3)
Improved mental health inventory	[55]	1 (3)
Decreased intracranial hemorrhage	[53]	1 (3)
Not reported	[22-26,28,29,31,34,36-39,41,42,44,47,49,51,52,56,57,61-66]	28 (N/A[b])

[a]eGFR: estimated glomerular filtration rate.

[b]N/A: not applicable.

Interactions Between Observations

Interventions of mHealth resulted in seven occurrences of a result (clinical and administrative outcomes) and six occurrences of an improvement in at least one clinical outcome. The interventions with telephone or televideo resulted in four instances of improved patient satisfaction and a decrease in eGFR and weight loss. The interventions of eHealth resulted in very few instances of either clinical or administrative outcomes other than improved compliance and health behaviors.

Discussion

Principal Findings

Telemedicine is examined in countries worldwide, and it is clear that the COVID-19 pandemic caused a rapid adoption of this modality of medicine to ensure the viability of practices. A key issue for discussion is the differences in findings between this systematic review and another recent similar review [14]. This systematic review identified key facilitators and barriers, and further analyzed health outcomes. The other similar review identified themes of effectiveness but failed to meet the expectations for a systematic review in terms of medical outcomes [68]. Common themes between the two reviews were: rapid telemedicine expansion, education, improved access, convenience, and patient satisfaction.

Summary of Evidence

This systematic review exercised a set Boolean search string in four common research databases to analyze 46 articles originating from five continents for themes of facilitators, barriers, and medical outcomes. Nearly 50% of the articles demonstrated the strongest evidence and nearly 60% demonstrated the highest quality of evidence. Various forms of telemedicine were examined: eHealth, mHealth, audio only, telemonitoring, telecoaching, telerehab, robotics or artificial intelligence, and televideo. Twenty-five facilitator themes and individual observations, 12 results themes and observations, and 7 medical outcome themes and observations were recorded and analyzed. Forty-one percent of barrier themes recorded either an improvement or no statistically significant improvement in results compared to the control group. Forty percent of the observations recorded an improvement in at least one medical outcome.

Health care administrators can focus on the findings demonstrating that implementation of telemedicine will increase self-management [27,30,35,37,38,40,41,45,46,49,57,58], adherence [32,34,35,37,38,40,43,46,58,67], access [24,27-30,41,52,53,60], and social support [49,50,54,55,62-64,66]. Telemedicine is shown to be an effective modality of treatment [28,32,35,38,40,41,43,46,48-50,53-55,58-60,64] at a decreased cost [31,46,52,56,59,60]. Patients perceive the modality to be convenient and easy to use [26,34,36,43-45,47], and its implementation increases patient satisfaction [24,25,43,47,49,50,63].

Health policymakers should focus on several barriers to increase the adoption of telemedicine. Because technical literacy, availability of technology, and connectivity are listed as the most often cited barriers, public programs should be offered to

assist those with these difficulties. Technical literacy is often associated along age or socioeconomic lines, and researchers acknowledge the dearth of research in the area of how to overcome this obstacle [69]. However, community centers that provide access to computers, classes on computers, and a dedicated broadband connection can all contribute to solutions to these barriers.

A key similarity between the 2020 systematic review [14] and this review is the rapid expansion of telemedicine. Eleven articles analyzed in this review used a phrase similar to "the pandemic created an acceptance of telemedicine technology" [23,25,31,33,41,43,47,51,52,61,62]. A systematic review published in 2018 cited cost as the chief barrier to adoption, whereas this review only found cost as a barrier in 8% of all observations [5]. The COVID-19 pandemic forced acceptance of the technology and enabled providers to not focus so intently on the cost of its implementation.

Limitations

This systematic review selected 46 articles for analysis from four commonly available research databases. A larger group for analysis could have yielded richer results. This review also only utilized two researchers to analyze the data; additional researchers could have identified additional themes. Selection bias was controlled through independent analysis of all articles by both reviewers followed by consensus meetings. Publication bias is the largest limitation because we were unable to query and analyze unpublished articles.

Conclusion

The COVID-19 pandemic caused huge problems to deliver medicine traditionally. However, these problems created an environment that limited face-to-face medical encounters and fostered legislation to reimburse the telemedicine modality for broad and rapid adoption of telemedicine to expand the access of care beyond the physical walls of the clinic. Physicians should feel confident that the telemedicine modality will be reimbursed and will have very little effect on patient satisfaction. Health care administrators who have not already adopted telemedicine should feel confident in the technology; however, they should ensure that sufficient confidentiality and security measures are in place. Policymakers should enact legislation to remove or mitigate barriers such as availability of technology, technical literacy, and connectivity, as these are commonly referred to in the literature.

Authors' Contributions

KH and CK contributed equally to this project. Both authors were involved in each step of the research and publication process.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sample size, bias, effect size, country of origin, and statistics used.
[DOCX File , 60 KB - jmir_v24i1e31752_app1.docx]

Multimedia Appendix 2
Observation-to-theme conversion for patient satisfaction as well as facilitators and barriers to adoption.
[DOCX File , 81 KB - jmir_v24i1e31752_app2.docx]

References

1. World Health Organization. Telemedicine: opportunities and developments in member states: report on the second global survey on eHealth 2009. Geneva, Swizerland: World Health Organization; 2010.
2. FitzGerald M, Williams R. Commonwealth Fund. 2020 Nov 20. URL: https://www.commonwealthfund.org/blog/2020/variations-telemedicine-across-world-qa-tiago-cravo-oliveira-hashiguchi [accessed 2021-02-15]
3. Bokolo AJ. Exploring the adoption of telemedicine and virtual software for care of outpatients during and after COVID-19 pandemic. Ir J Med Sci 2021 Feb;190(1):1-10 [FREE Full text] [doi: 10.1007/s11845-020-02299-z] [Medline: 32642981]
4. Fieux M, Duret S, Bawazeer N, Denoix L, Zaouche S, Tringali S. Telemedicine for ENT: Effect on quality of care during Covid-19 pandemic. Eur Ann Otorhinolaryngol Head Neck Dis 2020 Sep;137(4):257-261 [FREE Full text] [doi: 10.1016/j.anorl.2020.06.014] [Medline: 32624390]
5. Kruse SC, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2016 Oct 16;24(1):4-12. [doi: 10.1177/1357633x16674087]
6. Kuroda N. Decision making on telemedicine for patients with epilepsy during the coronavirus disease 2019 (COVID-19) crisis. Front Neurol 2020;11:722. [doi: 10.3389/fneur.2020.00722] [Medline: 32714276]
7. Ben-Pazi H, Beni-Adani L, Lamdan R. Accelerating telemedicine for cerebral palsy during the COVID-19 pandemic and beyond. Front Neurol 2020;11:746. [doi: 10.3389/fneur.2020.00746] [Medline: 32670193]
8. Royce TJ, Sanoff HK, Rewari A. Telemedicine for cancer care in the time of COVID-19. JAMA Oncol 2020 Nov 01;6(11):1698-1699. [doi: 10.1001/jamaoncol.2020.2684] [Medline: 32672821]

9. Hong Z, Li N, Li D, Li J, Li B, Xiong W, et al. Telemedicine during the COVID-19 pandemic: experiences from Western China. J Med Internet Res 2020 May 08;22(5):e19577 [FREE Full text] [doi: 10.2196/19577] [Medline: 32349962]
10. Mojdehbakhsh RP, Rose S, Peterson M, Rice L, Spencer R. A quality improvement pathway to rapidly increase telemedicine services in a gynecologic oncology clinic during the COVID-19 pandemic with patient satisfaction scores and environmental impact. Gynecol Oncol Rep 2021 May;36:100708 [FREE Full text] [doi: 10.1016/j.gore.2021.100708] [Medline: 33521218]
11. Chang PJ, Jay GM, Kalpakjian C, Andrews C, Smith S. Patient and provider-reported satisfaction of cancer rehabilitation telemedicine visits during the COVID-19 pandemic. PM R 2021 Dec;13(12):1362-1368 [FREE Full text] [doi: 10.1002/pmrj.12552] [Medline: 33455066]
12. Duarte A, Gouveia E Melo R, Lopes A, Rato JP, Valente J, Pedro LM. Lessons learned from the impact of the COVID-19 pandemic in a vascular surgery department and preparation for future outbreaks. Ann Vasc Surg 2021 May;73:97-106 [FREE Full text] [doi: 10.1016/j.avsg.2021.01.060] [Medline: 33493593]
13. Regelmann MO, Conroy R, Gourgari E, Gupta A, Guttmann-Bauman I, Heksch R, PES Drug and Therapeutics Committee and Practice Management Committee. Pediatric endocrinology in the time of COVID-19: considerations for the rapid implementation of telemedicine and management of pediatric endocrine conditions. Horm Res Paediatr 2020;93(6):343-350 [FREE Full text] [doi: 10.1159/000513060] [Medline: 33486483]
14. Betancourt JA, Rosenberg MA, Zevallos A, Brown JR, Mileski M. The impact of COVID-19 on telemedicine utilization across multiple service lines in the United States. Healthcare (Basel) 2020 Oct 01;8(4):380 [FREE Full text] [doi: 10.3390/healthcare8040380] [Medline: 33019667]
15. Kruse CS. Writing a systematic review for publication in a health-related degree program. JMIR Res Protoc 2019 Oct 14;8(10):e15490 [FREE Full text] [doi: 10.2196/15490] [Medline: 31527018]
16. Page M, McKenzie J, Bossuyt P, Boutron I, Hoffmann T, Mulrow C, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021 Mar 29;372:n71 [FREE Full text] [doi: 10.1136/bmj.n71] [Medline: 33782057]
17. Newhouse R, Dearholt S, Poe S, Pugh L, White K. The Johns Hopkins nursing evidence-based practice rating scale. Johns Hopkins University School of Nursing. 2005. URL: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjuheqN3-30AhXummoFHa7zBSsQFnoECAMQAQ&url=http%3A%2F%2Fevidencebasednurse.weebly.com%2Fuploads%2F4%2F2%2F0%2F8%2F42081989%2Fjhnedp_evidence_rating_scale.pdf&usg=AOvVaw0TqXuL2_eg1w4Da9XXoKNu [accessed 2021-02-15]
18. Light RJ. Measures of response agreement for qualitative data: Some generalizations and alternatives. Psychol Bull 1971;76(5):365-377. [doi: 10.1037/h0031643]
19. Braun V, Clarke V. Using thematic analysis in psychology. Qual Res Psychol 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
20. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
21. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22(3):276-282 [FREE Full text] [Medline: 23092060]
22. Ben-Arye E, Gressel O, Ben-Arye E, Samuels N. Feasibility of an online integrative oncology treatment program during COVID-19. J Pain Symptom Manage 2021 Feb;61(2):e1-e3 [FREE Full text] [doi: 10.1016/j.jpainsymman.2020.11.009] [Medline: 33189855]
23. Kurihara K, Nagaki K, Inoue K, Yamamoto S, Mishima T, Fujioka S, et al. Attitudes toward telemedicine of patients with Parkinson's disease during the COVID‐19 pandemic. Neurol Clin Neurosci 2020 Nov 18;9(1):77-82. [doi: 10.1111/ncn3.12465]
24. Richards AE, Curley K, Christel L, Zhang N, Kouloumberis P, Kalani MA, et al. Patient satisfaction with telehealth in neurosurgery outpatient clinic during COVID-19 pandemic. Interdisciplin Neurosurg 2021 Mar;23:101017. [doi: 10.1016/j.inat.2020.101017]
25. Yu J, Afridi SM, Cozart AC, Isea L, Guan J. Evaluation and feedback for telehealth from patients and physicians during the early stage of COVID-19 pandemic period. Cureus 2021 Jan 11;13(1):e12633 [FREE Full text] [doi: 10.7759/cureus.12633] [Medline: 33585121]
26. Alkire (née Nasr) L, O'Connor GE, Myrden S, Köcher S. Patient experience in the digital age: An investigation into the effect of generational cohorts. J Retail Consum Serv 2020 Nov;57:102221. [doi: 10.1016/j.jretconser.2020.102221]
27. Ballin M, Hult A, Björk S, Lundberg E, Nordström P, Nordström A. Web-based exercise versus supervised exercise for decreasing visceral adipose tissue in older adults with central obesity: a randomized controlled trial. BMC Geriatr 2020 May 12;20(1):173 [FREE Full text] [doi: 10.1186/s12877-020-01577-w] [Medline: 32398024]
28. Banbury A, Nancarrow S, Dart J, Gray L, Dodson S, Osborne R, et al. Adding value to remote monitoring: Co-design of a health literacy intervention for older people with chronic disease delivered by telehealth - The telehealth literacy project. Patient Educ Couns 2020 Mar;103(3):597-606. [doi: 10.1016/j.pec.2019.10.005] [Medline: 31744701]

29. Barnett A, Savic M, Pienaar K, Carter A, Warren N, Sandral E, et al. Enacting 'more-than-human' care: Clients' and counsellors' views on the multiple affordances of chatbots in alcohol and other drug counselling. Int J Drug Policy 2021 Aug;94:102910 [FREE Full text] [doi: 10.1016/j.drugpo.2020.102910] [Medline: 33059955]
30. Batalik L, Dosbaba F, Hartman M, Batalikova K, Spinar J. Benefits and effectiveness of using a wrist heart rate monitor as a telerehabilitation device in cardiac patients: A randomized controlled trial. Medicine (Baltimore) 2020 Mar;99(11):e19556. [doi: 10.1097/MD.0000000000019556] [Medline: 32176113]
31. Beller HL, Rapp DE, Noona SW, Winkelman AJ, Zillioux JM, Smith RP, et al. Tele-urology during COVID-19: rapid implementation of remote video visits. Urol Pract 2020 Nov;7(6):442-447. [doi: 10.1097/upj.0000000000000167]
32. Bernabe-Ortiz A, Pauschardt J, Diez-Canseco F, Miranda JJ. Sustainability of mHealth effects on cardiometabolic risk factors: five-year results of a randomized clinical trial. J Med Internet Res 2020 Apr 21;22(4):e14595 [FREE Full text] [doi: 10.2196/14595] [Medline: 32314970]
33. Bilgrami Z, Abutaleb A, Chudy-Onwugaje K, Langenberg P, Regueiro M, Schwartz DA, et al. Effect of TELEmedicine for inflammatory bowel disease on patient activation and self-efficacy. Dig Dis Sci 2020 Jan;65(1):96-103 [FREE Full text] [doi: 10.1007/s10620-018-5433-5] [Medline: 30604373]
34. Broers ER, Kop WJ, Denollet J, Widdershoven J, Wetzels M, Ayoola I, et al. A personalized eHealth intervention for lifestyle changes in patients with cardiovascular disease: randomized controlled trial. J Med Internet Res 2020 May 22;22(5):e14570 [FREE Full text] [doi: 10.2196/14570] [Medline: 32441658]
35. Cho SMJ, Lee JH, Shim J, Yeom H, Lee SJ, Jeon YW, et al. Effect of smartphone-based lifestyle coaching app on community-dwelling population with moderate metabolic abnormalities: randomized controlled trial. J Med Internet Res 2020 Oct 09;22(10):e17435 [FREE Full text] [doi: 10.2196/17435] [Medline: 33034564]
36. Claes J, Cornelissen V, McDermott C, Moyna N, Pattyn N, Cornelis N, et al. Feasibility, acceptability, and clinical effectiveness of a technology-enabled cardiac rehabilitation platform (Physical Activity Toward Health-I): randomized controlled trial. J Med Internet Res 2020 Feb 04;22(2):e14221 [FREE Full text] [doi: 10.2196/14221] [Medline: 32014842]
37. Coorey G, Peiris D, Neubeck L, Redfern J. A realist evaluation approach to explaining the role of context in the impact of a complex eHealth intervention for improving prevention of cardiovascular disease. BMC Health Serv Res 2020 Aug 18;20(1):764 [FREE Full text] [doi: 10.1186/s12913-020-05597-5] [Medline: 32811480]
38. Ding H, Jayasena R, Chen SH, Maiorana A, Dowling A, Layland J, et al. The effects of telemonitoring on patient compliance with self-management recommendations and outcomes of the innovative telemonitoring enhanced care program for chronic heart failure: randomized controlled trial. J Med Internet Res 2020 Jul 08;22(7):e17559 [FREE Full text] [doi: 10.2196/17559] [Medline: 32673222]
39. Geramita E, DeVito Dabbs AJ, DiMartini A, Pilewski J, Switzer G, Posluszny D, et al. Impact of a mobile health intervention on long-term nonadherence after lung transplantation: follow-up after a randomized controlled trial. Transplantation 2020 Mar;104(3):640-651 [FREE Full text] [doi: 10.1097/TP.0000000000002872] [Medline: 31335759]
40. Gong K, Yan Y, Li Y, Du J, Wang J, Han Y, APP Study Group et al. Mobile health applications for the management of primary hypertension: A multicenter, randomized, controlled trial. Medicine (Baltimore) 2020 Apr;99(16):e19715. [doi: 10.1097/MD.0000000000019715] [Medline: 32311957]
41. Han JK, Al-Khatib SM, Albert CM. Changes in the digital health landscape in cardiac electrophysiology: a pre-and peri-pandemic COVID-19 era survey. Cardiovasc Dig Health J 2021 Feb;2(1):55-62. [doi: 10.1016/j.cvdhj.2020.12.001]
42. Harding R, Carrasco JM, Serrano-Pons J, Lemaire J, Namisango E, Luyirika E, et al. Design and evaluation of a novel mobile phone application to improve palliative home-care in resource-limited settings. J Pain Symptom Manage 2021 Jul;62(1):1-9. [doi: 10.1016/j.jpainsymman.2020.09.045] [Medline: 33246073]
43. Hsia BC, Singh AK, Njeze O, Cosar E, Mowrey WB, Feldman J, et al. Developing and evaluating ASTHMAXcel adventures: a novel gamified mobile application for pediatric patients with asthma. Ann Allergy Asthma Immunol 2020 Nov;125(5):581-588 [FREE Full text] [doi: 10.1016/j.anai.2020.07.018] [Medline: 32711031]
44. Hsieh P, Lai H. Exploring people's intentions to use the health passbook in self-management: An extension of the technology acceptance and health behavior theoretical perspectives in health literacy. Technol Forecast Soc Change 2020 Dec;161:120328. [doi: 10.1016/j.techfore.2020.120328]
45. Hutchesson MJ, Taylor R, Shrewsbury VA, Vincze L, Campbell LE, Callister R, et al. Be Health for Your Heart: a pilot randomized controlled trial evaluating a web-based behavioral intervention to improve the cardiovascular health of women with a history of preeclampsia. Int J Environ Res Public Health 2020 Aug 10;17(16):5779 [FREE Full text] [doi: 10.3390/ijerph17165779] [Medline: 32785044]
46. Jiménez-Marrero S, Yun S, Cainzos-Achirica M, Enjuanes C, Garay A, Farre N, et al. Impact of telemedicine on the clinical outcomes and healthcare costs of patients with chronic heart failure and mid-range or preserved ejection fraction managed in a multidisciplinary chronic heart failure programme: A sub-analysis of the iCOR randomized trial. J Telemed Telecare 2018 Sep 07;26(1-2):64-72. [doi: 10.1177/1357633x18796439]
47. Katt BM, Imbergamo C, Fletcher D, Aita D, Nakashian M, Kwok M, et al. Telehealth for upper extremity conditions: perceptions of the patient and provider. JAAOS Glob Res Rev 2020 Sep 10;4(9):e20.00127-e20.00113. [doi: 10.5435/jaaosglobal-d-20-00127]

48. Kobe E, Diamantidis C, Bosworth H, Davenport C, Oakes M, Alexopoulos AS, et al. Racial differences in the effectiveness of a multifactorial telehealth intervention to slow diabetic kidney disease. Med Care 2020 Nov;58(11):968-973 [FREE Full text] [doi: 10.1097/MLR.0000000000001387] [Medline: 32833935]
49. Lai B, Bond K, Kim Y, Barstow B, Jovanov E, Bickel CS. Exploring the uptake and implementation of tele-monitored home-exercise programmes in adults with Parkinson's disease: A mixed-methods pilot study. J Telemed Telecare 2018 Aug 22;26(1-2):53-63. [doi: 10.1177/1357633x18794315]
50. Lemelin A, Paré G, Bernard S, Godbout A. Demonstrated cost-effectiveness of a telehomecare program for gestational diabetes mellitus management. Diabetes Technol Ther 2020 Mar;22(3):195-202. [doi: 10.1089/dia.2019.0259] [Medline: 31603351]
51. Manning B, Harpole A, Harriott E, Postolowicz K, Norton E. Taking language samples home: feasibility, reliability, and validity of child language samples conducted remotely with video chat versus in-person. J Speech Lang Hear Res 2020 Dec 14;63(12):3982-3990 [FREE Full text] [doi: 10.1044/2020_JSLHR-20-00202] [Medline: 33186507]
52. Marques G, Drissi N, Díez de la Torre I, de Abajo BS, Ouhbi S. Impact of COVID-19 on the psychological health of university students in Spain and their attitudes toward Mobile mental health solutions. Int J Med Inform 2021 Mar;147:104369. [doi: 10.1016/j.ijmedinf.2020.104369] [Medline: 33388481]
53. Martins SC, Weiss G, Almeida AG, Brondani R, Carbonera LA, de Souza AC, et al. Validation of a smartphone application in the evaluation and treatment of acute stroke in a comprehensive stroke center. Stroke 2020 Jan;51(1):240-246. [doi: 10.1161/strokeaha.119.026727]
54. McGillicuddy JW, Chandler JL, Sox LR, Taber DJ. Exploratory analysis of the impact of an mHealth medication adherence intervention on tacrolimus trough concentration variability: post hoc results of a randomized controlled trial. Ann Pharmacother 2020 Dec;54(12):1185-1193 [FREE Full text] [doi: 10.1177/1060028020931806] [Medline: 32506922]
55. Mo Y, Wang H, Huang G, Chu M. Effectiveness of nurse-led program on mental health status and quality of life in patients with chronic heart failure. Medicine (Baltimore) 2020 Aug 14;99(33):e21746. [doi: 10.1097/MD.0000000000021746] [Medline: 32872064]
56. Mustonen E, Hörhammer I, Absetz P, Patja K, Lammintakanen J, Talja M, et al. Eight-year post-trial follow-up of health care and long-term care costs of tele-based health coaching. Health Serv Res 2020 Apr;55(2):211-217 [FREE Full text] [doi: 10.1111/1475-6773.13251] [Medline: 31884682]
57. O'Shea O, Woods C, McDermott L, Buys R, Cornelis N, Claes J, et al. A qualitative exploration of cardiovascular disease patients' views and experiences with an eHealth cardiac rehabilitation intervention: The PATHway Project. PLoS One 2020;15(7):e0235274 [FREE Full text] [doi: 10.1371/journal.pone.0235274] [Medline: 32628688]
58. Perri MG, Shankar MN, Daniels MJ, Durning PE, Ross KM, Limacher MC, et al. Effect of telehealth extended care for maintenance of weight loss in rural US communities: a randomized clinical trial. JAMA Netw Open 2020 Jun 01;3(6):e206764 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.6764] [Medline: 32539150]
59. Piera-Jiménez J, Winters M, Broers E, Valero-Bover D, Habibovic M, Widdershoven JWMG, et al. Changing the health behavior of patients with cardiovascular disease through an electronic health intervention in three different countries: cost-effectiveness study in the Do Cardiac Health: advanced new generation ecosystem (Do CHANGE) 2 randomized controlled trial. J Med Internet Res 2020 Jul 28;22(7):e17351 [FREE Full text] [doi: 10.2196/17351] [Medline: 32720908]
60. Press VG, Arora VM, Kelly CA, Carey KA, White SR, Wan W. Effectiveness of virtual vs in-person inhaler education for hospitalized patients with obstructive lung disease: a randomized clinical trial. JAMA Netw Open 2020 Jan 03;3(1):e1918205 [FREE Full text] [doi: 10.1001/jamanetworkopen.2019.18205] [Medline: 31899529]
61. Ramírez-Correa P, Ramírez-Rivas C, Alfaro-Pérez J, Melo-Mariano A. Telemedicine acceptance during the COVID-19 pandemic: an empirical example of robust consistent partial least squares path modeling. Symmetry 2020 Sep 25;12(10):1593. [doi: 10.3390/sym12101593]
62. Ronan P, Mian A, Carr SB, Madge SL, Lorenc A, Robinson N. Learning to breathe with Tai Chi online - qualitative data from a randomized controlled feasibility study of patients with cystic fibrosis. Eur J Integr Med 2020 Dec;40:101229 [FREE Full text] [doi: 10.1016/j.eujim.2020.101229] [Medline: 33106755]
63. Sacco G, Lléonart S, Simon R, Noublanche F, Annweiler C, TOVID Study Group. Communication technology preferences of hospitalized and institutionalized frail older adults during COVID-19 confinement: cross-sectional survey study. JMIR Mhealth Uhealth 2020 Sep 18;8(9):e21845 [FREE Full text] [doi: 10.2196/21845] [Medline: 32896832]
64. Scheerman JFM, Hamilton K, Sharif MO, Lindmark U, Pakpour AH. A theory-based intervention delivered by an online social media platform to promote oral health among Iranian adolescents: a cluster randomized controlled trial. Psychol Health 2020 Apr;35(4):449-466. [doi: 10.1080/08870446.2019.1673895] [Medline: 31621423]
65. Schrauben S, Appel L, Rivera E, Lora C, Lash J, Chen J, CRIC Study Investigators. Mobile health (mHealth) technology: assessment of availability, acceptability, and use in CKD. Am J Kidney Dis 2021 Jun;77(6):941-950.e1. [doi: 10.1053/j.ajkd.2020.10.013] [Medline: 33309860]
66. Shareef MA, Kumar V, Dwivedi YK, Kumar U, Akram MS, Raman R. A new health care system enabled by machine intelligence: Elderly people's trust or losing self control. Technol Forecast Soc Change 2021 Jan;162:120334. [doi: 10.1016/j.techfore.2020.120334]

67. van Dijk MR, Koster MPH, Oostingh EC, Willemsen SP, Steegers EAP, Steegers-Theunissen RPM. A mobile app lifestyle intervention to improve healthy nutrition in women before and during early pregnancy: single-center randomized controlled trial. J Med Internet Res 2020 May 15;22(5):e15773 [FREE Full text] [doi: 10.2196/15773] [Medline: 32412417]
68. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Nov 19;18(1):143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
69. Neumeyer X, Santos SC, Morris MH. Overcoming barriers to technology adoption when fostering entrepreneurship among the poor: the role of technology and digital literacy. IEEE Trans Eng Manage 2021 Dec;68(6):1605-1618. [doi: 10.1109/TEM.2020.2989740]

Abbreviations

CARES: Coronavirus Aid, Relief, and Economic Security
eGFR: estimated glomerular filtration rate
JHNEBP: Johns Hopkins Nursing Evidence-Based Practice Rating Scale
MeSH: Medical Subject Headings
mHealth: mobile health
PICOS: participants, intervention, results, outcomes, study design
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 02.07.21; peer-reviewed by G Deckard, T Menser; comments to author 23.07.21; revised version received 23.08.21; accepted 22.11.21; published 04.01.22.

Please cite as:
Kruse C, Heinemann K
Facilitators and Barriers to the Adoption of Telemedicine During the First Year of COVID-19: Systematic Review
J Med Internet Res 2022;24(1):e31752
URL: https://www.jmir.org/2022/1/e31752
doi:10.2196/31752
PMID:34854815

Review

Effectiveness of Digital Counseling Environments on Anxiety, Depression, and Adherence to Treatment Among Patients Who Are Chronically Ill: Systematic Review

Karoliina Paalimäki-Paakki[1,2], MSci; Mari Virtanen[3], PhD; Anja Henner[2], PhD; Miika T Nieminen[4,5,6], PhD; Maria Kääriäinen[1,4,7], PhD

[1]Research Unit of Nursing Science and Health Management, University of Oulu, Oulu, Finland
[2]Degree Programme of Radiography and Radiation Therapy, Oulu University of Applied Sciences, Oulu, Finland
[3]School of Rehabilitation and Examination, Helsinki Metropolia University of Applied Sciences, Helsinki, Finland
[4]Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
[5]Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Oulu, Finland
[6]Department of Diagnostic Radiology, Oulu University Hospital, Oulu, Finland
[7]Oulu University Hospital, Oulu, Finland

Corresponding Author:
Karoliina Paalimäki-Paakki, MSci
Research Unit of Nursing Science and Health Management
University of Oulu
Aapistie 5 A
Oulu, 90220
Finland
Phone: 358 505365959
Email: karoliina.paalimaki-paakki@oamk.fi

Abstract

Background: Patients who are chronically ill need novel patient counseling methods to support their self-care at different stages of the disease. At present, knowledge of how effective digital counseling is at managing patients' anxiety, depression, and adherence to treatment seems to be fragmented, and the development of digital counseling will require a more comprehensive view of this subset of interventions.

Objective: This study aims to identify and synthesize the best available evidence on the effectiveness of digital counseling environments at improving anxiety, depression, and adherence to treatment among patients who are chronically ill.

Methods: Systematic searches of the EBSCO (CINAHL), PubMed, Scopus, and Web of Science databases were conducted in May 2019 and complemented in October 2020. The review considered studies that included adult patients aged ≥18 years with chronic diseases; interventions evaluating digital (mobile, web-based, and ubiquitous) counseling interventions; and anxiety, depression, and adherence to treatment, including clinical indicators related to adherence to treatment, as outcomes. Methodological quality was assessed using the standardized Joanna Briggs Institute critical appraisal tool for randomized controlled trials or quasi-experimental studies. As a meta-analysis could not be conducted because of considerable heterogeneity in the reported outcomes, narrative synthesis was used to synthesize the results.

Results: Of the 2056 records screened, 20 (0.97%) randomized controlled trials, 4 (0.19%) pilot randomized controlled trials, and 2 (0.09%) quasi-experimental studies were included. Among the 26 included studies, 10 (38%) digital, web-based interventions yielded significantly positive effects on anxiety, depression, adherence to treatment, and the clinical indicators related to adherence to treatment, and another 18 (69%) studies reported positive, albeit statistically nonsignificant, changes among patients who were chronically ill. The results indicate that an effective digital counseling environment comprises high-quality educational materials that are enriched with multimedia elements and activities that engage the participant in self-care. Because of the methodological heterogeneity of the included studies, it is impossible to determine which type of digital intervention is the most effective for managing anxiety, depression, and adherence to treatment.

Conclusions: This study provides compelling evidence that digital, web-based counseling environments for patients who are chronically ill are more effective than, or at least comparable to, standard counseling methods; this suggests that digital environments could complement standard counseling.

(J Med Internet Res 2022;24(1):e30077) doi:10.2196/30077

KEYWORDS

mHealth; mobile health; eHealth; digital health; mobile apps; smartphone apps; web-based; telemedicine; chronic diseases; noncommunicable diseases; web-based interventions; mobile phone

Introduction

Background

Chronic diseases account for 71% of all deaths globally. Furthermore, the recent rapid increase in the number of patients who are chronically ill will heavily burden the health care sector. This review focuses on the use of digital counseling environments among patients with cancer and cardiovascular, musculoskeletal, and colorectal diseases.

Patients who are chronically ill need a variety of counseling approaches at different stages of the disease. Patient counseling, which refers to the interaction between a patient and health care professionals, can strongly support the patient's sense of responsibility in adhering to their treatment [1]. Most of the novel patient counseling methods, for example, mobile, digital, or ubiquitous counseling, can increase adherence to treatment, which has been worryingly low among patients who are chronically ill [2]. Digital counseling environments can provide peer support through interaction; motivate self-care; and offer understandable, reliable, and up-to-date information to help patients better understand their disease and make lifestyle changes [3-5]. In addition, novel counseling methods can help manage patients' anxiety and fear, as well as enhance patient safety [2,6,7]. Nevertheless, the current knowledge base regarding digital counseling for anxiety and adherence to treatment among patients who are chronically ill seems to be fragmented, which highlights the need for a comprehensive summary of the available counseling approaches.

Digital, customer-oriented services may improve a patient's quality of life and functionality when the service is accessible regardless of place or time and tailored to the patient's specific needs [8,9]. Various technologies now enable the provision of such services, which can provide individual counseling to patients at the correct time and in an appropriate manner [10-12]. The provision of materials in different formats promotes tailored counseling approaches [11-15], with previous research demonstrating that patients value inclusivity, comprehensibility, availability, and flexibility in these services [4,13,16-18].

In recent years, digitalization has offered numerous opportunities for providing health care through digital channels. The World Health Organization defines digital health as "a broad umbrella term encompassing e-health, as well as developing areas such as the use of advanced computer sciences." Mobile health (mHealth) is a subarea of digital health, and is described as "the use of wireless mobile technologies for health," whereas another subarea, ubiquitous health, is defined as services delivered through ubiquitous technologies such as tags, sensors, and biometric devices [19]. The main objective of digital health could be described as using various technologies to support the achievement of health goals through the internet. However, the realm of digital health is wide and, as such, various terms have been applied in digital health research. This review focuses on web-based solutions and mobile apps that integrate knowledge sharing to create participative elements for the patient. Studies focusing on SMS text messaging and gaming were excluded.

According to the World Health Organization, digital and mobile technologies support health care systems through targeted and untargeted patient communication, patient-to-patient communication, personal health tracking, and citizen-based reporting. An important objective of digital health interventions is the widespread promotion of positive changes in behavior to prevent the onset of chronic disease.

The impacts of various digital health interventions on the management of chronic diseases, especially diabetes mellitus [20-26], cardiovascular diseases, and cancer, have been studied extensively by systematic reviews during recent years [10,27-47]. However, many of the studies that have been reviewed suffer from methodological shortcomings, that is, insufficient power to detect changes in outcomes and relatively short study duration [31]. This has led to a fragmented knowledge base, and a comprehensive description of the available digital health solutions—along with their effectiveness—is needed to further develop counseling for patients who are chronically ill. Research implications for future eHealth studies have recently been identified, categorized, and prioritized [48]. For example, randomized controlled trial (RCT) studies with large sample sizes and long follow-up periods, along with investigations of the cost-effectiveness and user acceptance of eHealth interventions, should be conducted in the near future. Furthermore, decision-makers will need an improved understanding of which components of the studied interventions, for example, frequency, duration and delivery mode, or patient characteristics, contribute most to the overall e ectiveness of an eHealth intervention. Ethical aspects, intervention safety, and translation of findings into a practical context were also identified as necessary research elements [48].

Objectives

This systematic review aims to identify and synthesize the best available evidence on the effectiveness of digital counseling environments for managing anxiety, depression, and adherence to treatment among patients who are chronically ill.

This review answers the following question: What is the effectiveness of the digital counseling environments compared

with control (eg, usual care) on anxiety and depression and clinical outcomes related to adherence to treatment?

Methods

Systematic Review

A systematic review of RCTs was conducted according to the Centre for Reviews and Dissemination and Joanna Briggs Institute guidelines [49,50]. The research adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [51] regarding the reporting of evidence.

Inclusion and Exclusion Criteria

The selection of studies was based on predefined inclusion and exclusion criteria, which are reported in the patient, intervention, comparison, and outcome format [50]. The review considered studies that included participants who were adult patients aged ≥18 years with chronic diseases; described interventions that were digital (mobile, web-based, or ubiquitous) counseling approaches; and reported outcomes that were patient outcomes (primary outcomes), that is, anxiety, depression, or adherence to treatment, and clinical indicators (secondary outcomes) related to adherence to treatment. The comparator was no treatment, standard care, or another type of intervention. All RCTs and quasi-experimental studies published in English, Finnish, or Swedish from 2008 to 2020 were considered; this specific time period was selected to reflect the growth and adoption of digital technologies.

Studies focusing on patients aged <18 years or describing patients with psychiatric disorders or substance abuse problems were excluded. Furthermore, studies focusing on traditional counseling, SMS text message counseling, or eHealth game development were excluded. Studies were also excluded if they measured any outcomes other than those defined in the inclusion criteria or were protocols, reviews, editorial papers, discussions, recommendations, or parts of books.

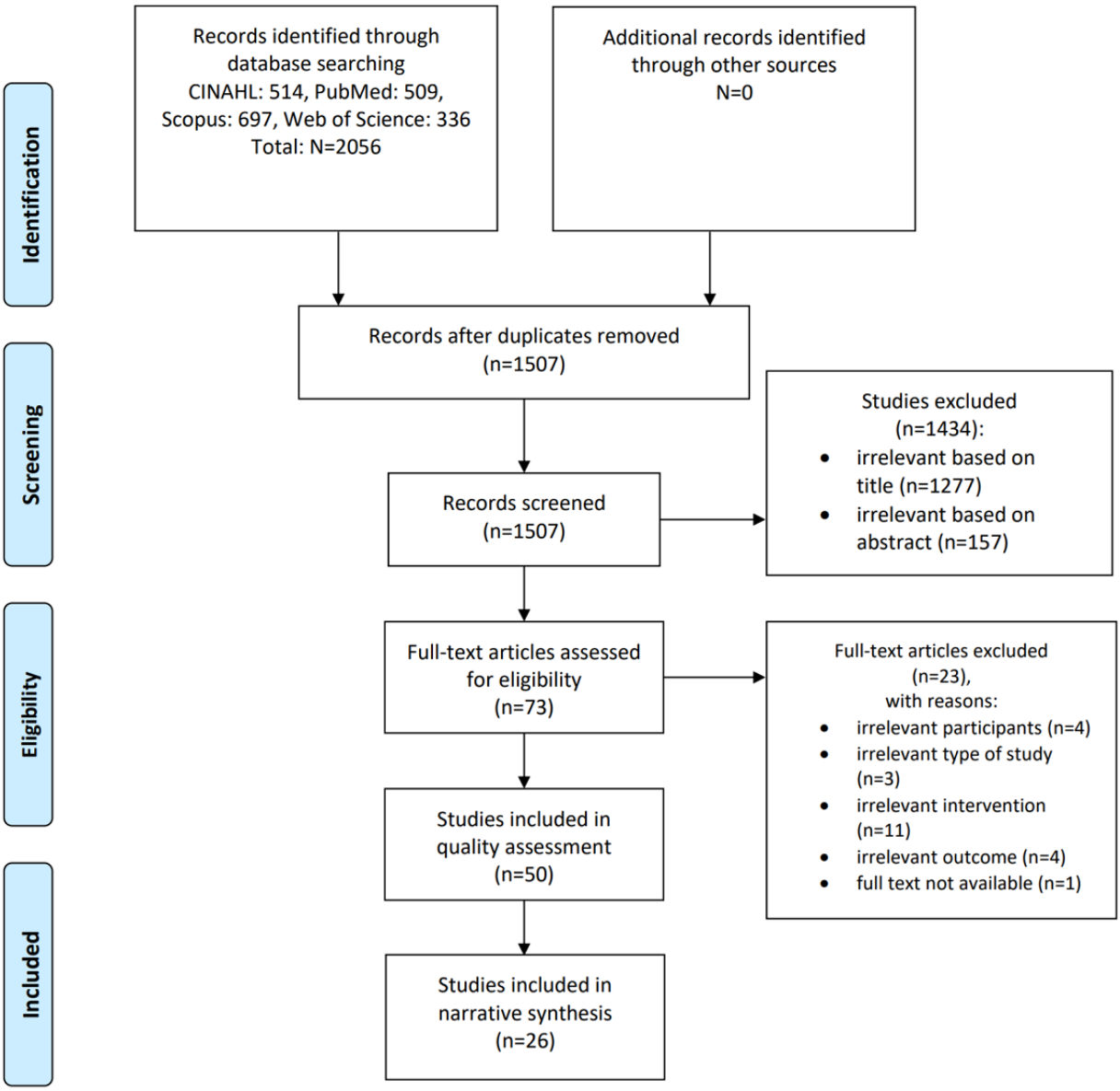
Search Strategy

Systematic literature searches were conducted across 4 electronic databases (CINAHL, PubMed, Scopus, and Web of Science) in May 2019, after which the search was complemented for the years 2019-2020 in October 2020. The reference lists of the included studies were screened for studies that may be relevant to the study objective, yet were not identified during the systematic literature search. An information specialist assisted the researchers in forming a search strategy and conducting the literature search. The search strategy for different databases is presented in Multimedia Appendix 1.

Study Selection

A total of 2056 publications were retrieved during the database searches. These publications were then imported into Zotero reference manager software (Corporation for Digital Scholarship). From the 2056 publications, 549 (26.7%) duplicates were removed. Of the 1507 studies remaining, 1434 (95.16%) were excluded after title and abstract screening by 2 independent researchers (KPP and MK) using predefined inclusion criteria, leaving 73 (4.84%) full-text articles relevant to the study objectives. Minor disagreements between the reviewers were resolved, and the researchers reached agreement. At the completion of the screening process, of the 73 studies, 50 (68%) met the inclusion criteria and were included in the critical appraisal. A PRISMA flow diagram was used to present the study selection process (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) 2009 flow diagram of study selection process. CINAHL: Cumulative Index to Nursing and Allied Health Literature.



Critical Appraisal

The methodological quality of the 50 selected studies was independently assessed by 2 researchers (KPP and MK) using a standardized Joanna Briggs Institute critical appraisal tool for RCTs and quasi-experimental studies [52]. The methodological quality was evaluated by assigning points to each criterion of the appraisal tool. Studies that scored at least 60% (8/13 points for RCTs and 5/9 points for quasi-experimental studies) across the appraisal criteria were included in the review. Of the 50 selected studies, 26 (52%) were included in the final review, whereas 24 (48%) were excluded based on poor blinding, unreliable measurement of outcomes, or inappropriate statistical analysis. Critical appraisal of the selected randomized controlled trial studies is presented in Multimedia Appendix 2 [53-76].

In the critical appraisal of the selected quasi-experimental studies (2/26, 8%), both scored 8 points out of 9. Each study had a control group, and there were no differences among the participants in the compared groups. Other than the intervention of interest, there were no differences between the groups in terms of care received. Multiple measurements of the outcomes both before and after the intervention were collected in the same way in both studies, and appropriate statistical analysis was conducted. The only unclear criterion concerned whether the follow-up was complete, and if not, whether differences between the groups in terms of their follow-up were adequately described and analyzed.

Data Extraction

Data from the original studies included in the review were extracted to meet the Centre for Reviews and Dissemination information requirements [49]. The first author (KPP) entered the extracted data into a standardized form (Multimedia Appendix 3 [53-78]) that also included the quality assessment scores. The second author (MV) confirmed the extracted data.

Because of the heterogeneity of outcomes reported in the identified RCTs, a meta-analysis was not possible [49]. Narrative synthesis was used to answer the research question.

Results

Characteristics of the Included Studies

A total of 26 studies were included in this review: 20 (77%) RCTs [53-72], 4 (15%) pilot RCTs [73-76], and 2 (8%) quasi-experimental studies [77,78] published from 2010 to 2020 in English in 13 countries. Details of the included studies are described in Multimedia Appendix 3.

Participants

The participants in the included studies were adult patients aged ≥18 years with a range of diseases: various cancers [53,63,65,66,69,70,73,75-77], along with cardiovascular [54,56-58,64,68,71,72,74], musculoskeletal [59,62], colorectal [55,60,61], and kidney diseases [78]. The sample size ranged from 29 to 1000 patients, with, of the 26 studies, 9 (35%) having enrolled fewer than 99 participants [58-60,62,64,68,73,76,78], 9 (35%) having enrolled 100-200 participants [53,54,61,66,69,70,74,75,77], 7 (27%) having enrolled 201-500 participants [55-57,63,65,67,72], and 1 (4%) having enrolled 1000 participants [71]. The follow-up period ranged from days to 12 months, more specifically, 3 months or less in 46% (12/26) of the studies [59-61,64,66,68-70,73,74,76,78], 3-6 months in 23% (6/12) of the studies [58,62,63,65,67,71], and 6-12 months in 31% (8/26) of the studies [53-57,72,75,77].

Interventions

Overview

The patient counseling environments described in the original publications included websites [53,55-57,59,62,63,65,66,72,75,76], mobile apps [54,60,61,67-71,73,74,77,78], or a combination of both [58,64]. Websites could be accessed with all devices, whereas the mobile apps were accessible with either a mobile phone or a tablet. The mobile apps were designed to be both iOS and Android compatible [54,61,64,74], only iOS compatible [58,60,67,73], or only Android compatible [68,77,78]. The interventions described in this review were heterogeneous, with detailed information provided in Multimedia Appendix 3. The interventions are described according to the Template for Intervention Description and Replication checklist [79].

Websites

Of the 26 studies, websites were the primary counseling approach in 12 (46%). Of these 12 studies, 6 (50%) focused on patients with cancer [53,63,65,66,75,76], 1 (8%) focused on patients with colorectal disease [55], 2 (17%) concerned patients with musculoskeletal disease [59,62], and 3 (25%) covered patients with cardiovascular disease [56,57,72]. The format, amount, and use of the counseling materials in the presented websites varied among the studies.

Counseling materials provided disease- or condition-specific information in different formats. The materials were gathered as learning material libraries [55,65,66,72,75,76], link collections [53,63,72,75], and patient stories [56,72,75].

In addition, 92% (11/12) of the presented websites included information-processing functionalities [53,55-57,59,62,63,65,66,72,76]. Participants were encouraged to assess, self-monitor, and report personal health data such as heart rate, blood pressure, blood glucose level, symptoms, distress, medication adherence, daily exercise, and diet [53,55,57,59,63,65,66,72,76] and fill in web-based medical, risk factor, and lifestyle forms [55-57,65,66]. Adherence to treatment was followed by learning tasks [56,66], an e-notebook or diary [53,55,62,63,66], and action plans [57,72,76]. Participants also received personalized advice, recommendations, and feedback based on their activity and self-reports [53,55-57,59,63,65,66]. All the websites described in 100% (12/12) of the studies included personalized content [53,55-57,59,62,63,65,66,72,75,76], that is, counseling, recommendations, and feedback based on the participants' inputs and responses.

Web-based patient–provider counseling was integrated into 58% (7/12) of the presented websites: web-based communication occurred through e-messages [53,55,57,65,66,75] and videoconferencing [75]. For 8% (1/12) of the websites, participants had the option to save an updated list of questions for the health care team [76]. Furthermore, 25% (3/12) of the studies included a component, that is, anonymous web-based forum group discussions [53,62,76] and blog [53], through which users could share their experiences with other patients.

Website use and activity were measured in 17% (2/12) of the included studies [53,63]; more specifically, these studies applied the following website analytics: total hits per user session [53], hits on program modules and pages [53], total viewing time [63], number of website log-ins per person [53,63], number of measures uploaded, amount of e-messages sent, and number of diary notes and posts in blog [53].

Mobile Apps

Of the 26 studies, mobile apps were the primary approach used in 12 (46%). Of these 12 studies, 5 (42%) focused on patients with cardiovascular diseases [54,67,68,71,74], 4 (33%) covered patients with cancer [69,70,73,77], 2 (17%) focused on patients with colorectal diseases [60,61], whereas 1 (8%) was conducted on patients requiring hemodialysis [78].

The format, amount, and use of the counseling materials in the mobile app varied among the studies. In addition to counseling materials, 67% (8/12) of the presented apps included information-processing functionalities [54,60,61,67,68,70,71,74]. Participants were encouraged to self-monitor and report health data such as blood pressure, physical activity, diet, medication adherence, symptoms, and sleep [54,60,67,68,70,71,78]. Participants would then receive personalized recommendations, feedback [54,60,67,68,70,71,74,78], and timed notifications [61,67,68,71,77,78] based on their activity and the information they entered into the app. Furthermore, of the 12 apps, 1 (8%) included a personal health record [74] that a patient could update with laboratory test results and use as a risk assessment tool.

The input data were also used as a clinical decision support tool by doctors [74]. The presented apps promoted adherence to treatment through daily or weekly challenges [54,60,67], a diary feature [54,78], and homework exercises [69]. Of the apps, 100% (12/12) provided personalized content [54,60,61,67-71,73,74,77,78], that is, counseling, recommendations, feedback, or notifications based on the participants' inputs and responses.

Of the 12 apps, web-based counseling was integrated into 3 (25%). Of these 3 apps, 1 (33%) included 60 minutes of individual counseling with a registered dietitian [54], 1 (33%) involved conversational messages that were responded to by artificial intelligence [67], and 1 (33%) included counseling through a bulletin board and SMS text messaging with the researcher [78].

App use was measured and reported in 33% (4/12) of the studies. For example, the patient satisfaction rate was calculated [61,71,74]; the frequency of app use was measured [70,71]; and the usability, feasibility, and acceptability of the app were investigated [74]. Of the 4 studies, 2 (50%) reported that the apps were rated as user friendly and easy to use, as well as helpful or indispensable [61,74], whereas only 15% of the participants in 1 (25%) study perceived the app to be very useful, with more than half of the participants perceiving the app to be of little use [71].

Combination of Website and Mobile App

Of the 26 studies, 2 (8%) concerning patients with cardiovascular disease used the combination of a website and mobile app as the primary approach for improving patients' mental health and adherence to treatment [58,64]. The format, amount, and use of the counseling materials varied between these 2 reports.

In addition to counseling materials, both studies reported that the presented intervention included information-processing functionalities. For example, participants used their mobile phone to enter health data such as blood pressure, blood glucose level, medication adherence, and diet [58,64]. Adherence to treatment was promoted by learning tasks and homework [64], action plans for lifestyle change [64], and reminders for self-monitoring [58]. Participants received automated, personalized recommendations and feedback based on their activity and input [64]. Furthermore, of the 2 interventions, 1 (50%) enabled the sharing of data, that is, a patient could share their personal health record with family members, caregivers, and health professionals [58]. Both the described interventions included personalized content [58,64].

Outcomes

Overview

Of the 26 studies included in this review, 13 (50%) measured anxiety and depression [53,55,56,62,63,65,66,69,70,73,74,76,77], 9 (35%) measured adherence to treatment [54,55,57,59-61,71,75,78], and 9 (35%) studied how the digital environment affected ≥1 clinical outcomes related to adherence to treatment [54,57,58,64,67,68,71,72,78]. The scales used to measure these outcomes are described in Textbox 1.

Textbox 1. The scales used to measure the outcomes.

Outcomes and scales

- Anxiety and depression
  - Hospital Anxiety and Depression Scale [53,55,62,63,65,66,69,70,73,76,77]
  - General Anxiety Disorder Scale-7 [56]
  - Hamilton Anxiety Rating Scale [69]
  - Distress Thermometer [66,76]
  - Impact of Events Scale [76]
  - EuroQol [74]
  - EuroQol 5-Dimensional Questionnaire, Youth Version [74]
  - Patient Health Questionnaire-9 [56,69]
- Adherence to treatment
  - Mediterranean Diet Score [54]
  - Compliance Questionnaire [55]
  - Medical record reviews [75]
  - Framingham Risk Score [57]
  - Numerical scale from 0 to 10 [59]
  - Mean adherence to predefined bundle of patient-dependent elements [60]
  - Compliance with the first low-fiber dietary change and duration of use of the clear liquid diet and bowel cleanliness using 3 scales: the modified Aronchick scale, the Ottawa Bowel Preparation Scale, and the Chicago Bowel Preparation Scale [61]
  - Morisky Medication Adherence Scale [71]
  - Compliance of Patient Role Behavior Tool [78]
- Clinical indicators related to adherence to treatment
  - Blood pressure [54,57,58,64,67,71,72]
  - Weight [54,57,64,78]
  - Total cholesterol [54,57,64,68]
  - High-density lipoprotein cholesterol [54,57,64,68]
  - Low-density lipoprotein cholesterol [54,57,64,68]
  - Triglycerides [54,64,68]
  - Glycosylated hemoglobin level [54,57]
  - High-sensitivity C-reactive protein [54,57]
  - Serum glucose values [64]
  - Frequency of alcohol consumption and frequency of smoking [57,58,64,71]
  - Frequency of exercise [57,58,64]
  - Exercise stress test [64]
  - Alanine aminotransferase, creatinine, and plasma carotenoids [57]

Anxiety and Depression

An analysis of the 26 identified studies revealed that 3 (12%) reported a statistically significant reduction in anxiety and depression. A few of the presented websites significantly reduced anxiety and depression among patients with cancer [53,65], and a mobile app decreased anxiety and depression among patients with cardiovascular disease [74] in comparison with the control groups.

In an RCT of web-based self-management support for 167 patients with breast cancer, the web choice group reported significantly lower anxiety (mean difference –0.79, 95% CI –1.49 to –0.09; *P*=.03) and depression (mean difference –0.79, 95% CI 1.18 to –0.05; *P*=.03) scores than the usual care group [53]. A web-based tailored program for 273 cancer survivors

was able to significantly decrease patients' Hospital Anxiety and Depression Scale (HADS) score (mean difference –0.90, intervention group SD 3.83-2.79 vs control group SD 3.86-2.59; 95% CI –1.51 to –0.29) compared with the control treatment [65]. In a pilot RCT that included 209 patients with atrial fibrillation [74], a mobile app reduced patients' anxiety and depression (P=.02) compared with the group that did not use the app.

Of the 26 studies, in 9 (35%), the experimental group exhibited positive changes in anxiety and depression; however, these changes were not statistically significant when compared with the results of the control group [55,56,62,63,66,69,70,76,77]. Of the 24 RCTs, 2 (8%) [63,66] assessed the effectiveness of an informational website in reducing distress among patients with cancer. Among 337 patients with breast cancer [63] and 129 patients newly diagnosed with cancer [66], the mean levels of anxiety or depression did not significantly differ between the intervention and control groups. However, the entire study population exhibited a significant decrease in the HADS score in 50% (1/2) of these studies (P=.03) [66]. According to the visual analog scale score (which ranges from 0 to 10), the intervention group showed significantly lower levels of distress than the control group (mean difference –0.85, 95% CI –1.60 to –0.10; P=.03) 2 months after an intervention [66].

In a pilot RCT [76] for patients with advanced ovarian cancer, no differences between the intervention and control groups were observed for any distress measure, although the group using the patient-centered, information-based website demonstrated lower, albeit nonsignificant, general distress as measured by the Distress Thermometer. In a double-center study of patients with mild to moderate ulcerative colitis [55], the patients in the control group in Denmark showed a significant improvement in depression (P=.01) compared with those in the intervention group, whereas the patients in Ireland who had used the tested website demonstrated a significant improvement in anxiety (P=.02) compared with those in the control group [55]. Of the 24 RCTs, 2 (8%) studies, with 1 (50%) that included patients with implantable cardioverter defibrillators [56] and 1 (50%) that included patients undergoing lumbar spine fusion [62], found that a web-based platform for anxiety did not significantly affect patients' anxiety and depression. Furthermore, RCTs investigating the effect of mobile apps on anxiety in patients with incurable cancer [69] and patients undergoing breast cancer chemotherapy [70] reported that both study groups experienced improvements in anxiety and depression, but no significant between-group differences existed. However, subsequent analyses of a subgroup of patients with severe baseline anxiety revealed that patients using the tested app showed greater improvements in the Hamilton Anxiety Rating Scale score (mean difference 7.44, SE 3.35; P=.04) and the HADS score (mean difference 4.44, SE 1.60; P=.01) than those in the control group [69]. A Taiwanese quasi-experimental study reported that a web-based intervention did not significantly improve distress, anxiety, and depression among breast cancer survivors [77]. A pilot RCT study of female patients undergoing surgery for breast cancer [73] reported similar between-group anxiety and depression scores both preoperatively and immediately after surgery; however, the control group, which did not have access to the additional information provided by the mobile app, showed significantly lower anxiety and depression scores (P=.03 and P=.02, respectively) 7 days after surgery.

Adherence to Treatment

Of the 26 studies, 4 (15%)—2 (50%) of which tested a website [55,59] and 2 (50%) of which presented a mobile app [74,78]—reported statistically significant improvements in adherence to treatment in the intervention group compared with the control group. Lambert et al [59] evaluated the effect of a home exercise website with remote support on self-reported exercise adherence among 80 people with upper or lower limb musculoskeletal conditions. The mean between-group difference for self-reported exercise adherence was 1.3 (11 points; 95% CI 0.2-2.3) in favor of the intervention group, which was a statistically significant result (P=.01). A double-center RCT in Denmark and Ireland reported better ulcerative colitis compliance among Danish patients who had used the tested websites than among those in the control group after 12 months, with adherence to 4 weeks of acute treatment also significantly better among the patients in the intervention group (73% compared with 42% among patients in the control group; P=.005) [55]. At the Irish center, the patients in the intervention group also showed better adherence to 4 weeks of acute treatment than those in the control group (73% vs 29%; P=.03). Moreover, a mobile app for patients with atrial fibrillation significantly improved drug adherence (P<.001) and anticoagulant satisfaction (P=.01) compared with usual care [74]. In a quasi-experimental study of self-management among 84 patients requiring hemodialysis, the use of a mobile app significantly improved self-efficacy compared with the results from patients in the control group (mean 4.79, SD 3.51 vs mean −1.05, SD 2.05; t_{82}=−9.30; P<.001). Treatment compliance also significantly increased in the experimental group (mean 11.57, SD 7.63) compared with the control group (mean −1.74, SD 2.71; t_{82}=−10.66; P=.001) [78].

Of the 26 studies included in this review, 6 (23%) found no significant between-group differences in adherence to treatment, and 2 (8%) evaluated how the use of a mobile app affects adherence to care among patients with cardiovascular disease [54,71]. The presented asynchronous dietary counseling mobile app resulted in a significantly larger proportion of participants who complied with the Mediterranean diet (Mediterranean Diet Scale score ≥9) over time (P<.001); however, no significant between-group differences were discerned. An RCT focusing on a mobile app for patients who had undergone surgical coronary revascularization did not reveal any significant between-group differences in mean medication adherence scores (mean difference 0.052, 95% CI –0.087 to 0.191; P=.46) at the 6-month follow-up point [71]. Keyserling et al [57] investigated whether a web-based lifestyle (n=193) and a medication intervention (n=192) can reduce coronary heart disease risk. Both intervention formats reduced coronary heart disease risk through the 12-month follow-up period; however, no significant between-group differences were found [57].

Helzlsouer et al [75] reported that a web-based navigation program for newly diagnosed low-income patients with breast cancer did not significantly affect treatment completion

compared with the control group. No significant between-group differences in the assessed measures of adherence were observed in 20% (2/10) of the RCTs of mobile apps, the first of which aimed to improve adherence as part of a recovery program after colorectal surgery [60] and the second aiming to improve adherence to bowel cleanliness among patients who had undergone colonoscopy [61]. However, both the apps were rated as user friendly and a better alternative to paper instructions (P<.001). [61]

Clinical Outcomes Related to Adherence

Of the 26 studies, 3 (12%) concerning patients with cardiovascular disease—1 (33%) presented the combination of a smartphone app and a website [58], 1 (33%) studied the effectiveness of a website [72], and 1 (33%) presented a mobile app [68]—reported that the intervention group differed significantly from the control group in terms of clinical indicators related to adherence to treatment.

In a 6-month–long RCT that included 95 participants with hypertension, the combination of a smartphone app and website yielded significant improvements in clinical indicators related to adherence among patients in the intervention group compared with those in the control group [58]. More specifically, the results showed reduced consumption of cigarettes (P<.001) and decreased systolic and diastolic blood pressure levels (baseline: 140.6/89.4 mm Hg; end of study: 136.5/83.9 mm Hg) in the patients in the intervention group compared with those in the control group. Furthermore, the frequency at which the patients in the intervention group achieved blood pressure control increased from 45% to 59%. Similarly, e-counseling for patients with hypertension (n=264) resulted in a significant reduction in systolic blood pressure after 12 months in the patients in the intervention group compared with those in the control group (–10.1, 95% CI –12.5 to –7.6 mm Hg vs –6.0, 95% CI –8.5 to –3.5 mm Hg; P=.02) [72]. A 12-week smartphone app intervention for 57 patients with cardiovascular disease led to significant reductions in both triglyceride and total cholesterol levels in the intervention group compared with the control group (P=.02 and P=.01, respectively) [68]. In the same study, medication adherence also significantly increased in the intervention group (43.33% vs 82.14%; P=.002), whereas the control group only showed a minor increase (30% vs 37.93%; P=.52). This between-group difference was statistically significant (82.14% vs 37.93%; P=.001). No significant between-group changes were found with respect to low-density lipoprotein and high-density lipoprotein levels [68].

Digital health interventions for 80 patients with acute coronary syndrome [64], 100 patients with cardiovascular disease [54], and 84 patients requiring hemodialysis revealed improved weight loss in the intervention group compared with the control group (mean −5.1, SD 6.5 kg vs mean −0.8, SD 3.8 kg; P=.02 [64]; 1.5 kg vs 1.4 kg; P=.04 [54]; and mean −0.56, SD 0.88 vs mean 0.05, SD 1.08; P=.005 [78], respectively). Among the patients with cardiovascular disease, the digital health intervention did not significantly affect systolic blood pressure [54,57,64,67,71], diastolic blood pressure [54,64,71], lipids [54,57,64], blood glucose level [64], glycosylated hemoglobin level, C-reactive protein [54], or smoking frequency [57,71].

Discussion

Principal Findings

This systematic review identified and synthesized the best available evidence regarding how effective digital counseling interventions are at managing anxiety, depression, and adherence to treatment among patients who are chronically ill.

The results indicate that an effective digital counseling environment includes both high-quality educational material, possibly enriched with multimedia elements, and activities that engage participants. Because of the heterogeneity of the studies included in this review, it was impossible to determine which type of digital intervention was the most effective at managing anxiety, depression, and adherence to treatment. Furthermore, determining the aspects responsible for changes in self-management was difficult. Overall, digital, web-based counseling environments designed for patients who are chronically ill seem to be more effective than, or at least comparable to, standard counseling methods. This indicates that well-accepted digital environments could complement standard counseling. Patients should be afforded a variety of web-based educational resources that correspond to their care objectives and needs. These services should also be provided at an appropriate time to ensure maximum benefits [39]. Previous reviews have identified the highly participative features of mHealth interventions, for example, reminders and continuous feedback, patient-centeredness, individually tailored content, and patient–provider communication, to be a large advantage of these services [28,31,34,41,80,81]. Furthermore, it has previously been suggested that digital environments have the potential to increase patient involvement, empowerment, and security through increased knowledge, symptom management, participation, engagement, and improved clinician–patient communication [82,83]. These types of services also do not depend on location, which will improve access to care for patients in remote locations where other services may not be available and, therefore, counteract care inequity. In light of the COVID-19 pandemic, digital environments can also support patients who are chronically ill and living in isolated circumstances [9].

Digital counseling environments can enhance clinical practice and care by empowering patients with chronic disease self-management, reducing dependency on health care professionals, and possibly changing the chronic disease course in the long term. Furthermore, digital counseling environments can be accommodated and used for other patient groups by enhancing diagnostic examination success and optimizing care procedures.

Nevertheless, digital environments can also contribute to care inequity if certain patients do not have the ability or resources to access digital environments. Moreover, digital environments can cause ambivalence and uncertainty if patients lack the digital skills and knowledge of how to use these environments [82].

Surprisingly, all the studies included in this review were based on basic technologies, that is, internet-based environments, websites, and mobile apps. There were no reports of

interventions that applied emerging technologies such as augmented reality, virtual reality, mixed reality, or 360° virtual reality solutions. Furthermore, none of the presented digital counseling approaches used ubiquitous elements, for example, tags or sensors.

Effects of Digital Counseling Environment on Patient and Clinical Outcomes

Digital interventions significantly improved anxiety and depression among patients with cancer [53,65] and cardiovascular disease [74]. Positive, albeit statistically nonsignificant, changes in anxiety and depression were also measured among patients with cancer [63,66,69,70,76,77], as well as individuals with colorectal [55], cardiovascular [56], and musculoskeletal [62] diseases. However, a pilot study [73] found that patients in the control group—who did not have access to the additional information provided by the mobile app—showed significantly lower anxiety and depression than the intervention population. As this particular study explored patients with cancer, it is possible that the increased amount of knowledge in the app reminded women of the cancer treatment they were going through. In contrast, a systematic review [38] reported that 17 studies found eHealth solutions to improve anxiety among patients with breast cancer. Nevertheless, other studies have reported eHealth interventions to exert contradictory effects on anxiety [40,83]. This includes the surprising finding that increased knowledge does not necessarily reduce anxiety. This area of research clearly needs to be investigated in more detail.

Various digital counseling approaches significantly improved adherence to treatment among patients requiring hemodialysis [78], as well as individuals with musculoskeletal [59], colorectal [55], and cardiovascular [74] diseases. In 23% (6/26) of the studies, although adherence to treatment increased among patients with cardiovascular [54,57,71] and colorectal diseases [60,61] and cancer [75], no statistically significant differences between the groups (intervention and control) were found. Digital interventions also significantly improved the clinical indicators related to adherence to treatment among patients with cardiovascular disease [58,68,72]. For example, eHealth interventions significantly improved adherence to treatment [84] and blood pressure control [29,37,41,42,45,85]. Recent systematic reviews of mHealth interventions for hypertension [28] and coronary artery disease [30] have provided evidence that mHealth interventions are effective for blood pressure control, self-management, and medication adherence. It should be noted that the overall risk of bias was relatively high in both these studies.

The lack of significant improvements in the outcomes of patients who are chronically ill after digital counseling interventions may be explained by various methodological issues such as a short follow-up period or insufficient power to detect changes in outcomes [31,33,34,85]. A recent umbrella review [31] concluded that telemedicine has the potential to improve clinical outcomes in patients with diabetes; however, it was not found to have a significant and clinically meaningful impact on blood pressure because the outcomes measuring blood pressure showed low overall certainty. Risk of bias; inconsistency; differences in patient populations, settings, and interventions; imprecision; publication bias; and the underreporting of relevant information have been listed as the main reasons why previous reports have only provided low-quality evidence concerning the effectiveness of digital counseling approaches. In addition, the heterogeneity in eHealth definitions also makes between-study comparisons difficult, which are necessary to provide health care professionals with evidence-based recommendations [30,48,86,87].

Limitations

This review includes a few inherent limitations. The literature search conducted for this review excluded gray literature, which means that relevant studies may have been overlooked. Language limitations were not used during the search process, but only studies published in English, Finnish, and Swedish were considered during the screening process. This may have resulted in language bias.

A further limitation was the varying quality and heterogeneity of the selected studies, that is, sample sizes, type of interventions, and length of follow-up times, which differed among the studies. The sample sizes were small (fewer than 200 patients overall) in 65% (17/26) of the studies. The complex digital counseling interventions were diverse and heterogeneous in content and had various risks of bias in their methodology. Quality assessment was performed using a standardized Joanna Briggs Institute critical appraisal tool for RCTs and quasi-experimental studies [52] to avoid systematic bias. Of the 24 RCT studies, 9 (38%) scored at least 10 points out of 13, whereas 15 (62%) scored less than 10 points out of 13. Of the 24 RCT studies, 2 (8%) had the lowest score, 8 points out of 13. Both quasi-experimental studies were rated as good quality. The highest risk of bias in the selected studies related to blinding of participants and personnel, blinding of outcome assessment, and incomplete outcome data.

Of the 26 studies, 6 (23%) did not measure anxiety, depression, or adherence to treatment as a primary outcome of digital counseling interventions. In addition, several different scales were used to measure the selected primary outcomes. Because of the heterogeneity of the outcomes measured and scales used in the included studies, we could not perform a meta-analysis. This may have introduced additional bias in the results.

The review was strengthened by the use of a systematic and extensive search process that used several databases and was conducted with the assistance of an information specialist. To avoid subjective selection bias, studies were selected for inclusion by 2 researchers (KPP and MK) working independently.

Conclusions

Among the 26 included studies, 10 (38%) digital, web-based interventions demonstrated statistically significant positive effects on anxiety and depression, adherence to treatment, and clinical indicators related to adherence to treatment. Positive, albeit statistically nonsignificant, changes were reported in 69% (18/26) of the studies. These results indicate that digital environments may improve anxiety, depression, and adherence

to treatment among patients who are chronically ill, and hence have significant repercussions for the health care sector.

Acknowledgments

The authors thank Business Finland (grant 6557/31/2016) and the Scholarship Fund of the University of Oulu (grants 20180022 and 20210019) for financial support during the protocol, data collection, and publication phases. These institutions had no role in the findings or preparation of the manuscript.

Authors' Contributions

The study was designed by KPP, MV, AH, and MK. KPP was responsible for data collection, data analysis, and drafting the manuscript, whereas MV, AH, MN, and MK made critical and intellectual revisions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy for different databases.
[DOCX File , 14 KB - jmir_v24i1e30077_app1.docx]

Multimedia Appendix 2
Critical appraisal of the selected randomized controlled trial studies (N=24).
[DOCX File , 20 KB - jmir_v24i1e30077_app2.docx]

Multimedia Appendix 3
Details of the included studies.
[DOCX File , 54 KB - jmir_v24i1e30077_app3.docx]

References

1. Kääriäinen M. Potilasohjauksen Laatu: Hypoteettisen Mallin Kehittäminen. University of Oulu, Finland. 2007. URL: http://urn.fi/urn:isbn:9789514284984 [accessed 2020-09-28]
2. Kotseva K, Wood D, De Bacquer D, De Backer G, Rydén L, Jennings C, EUROASPIRE Investigators. EUROASPIRE IV: a European Society of Cardiology survey on the lifestyle, risk factor and therapeutic management of coronary patients from 24 European countries. Eur J Prev Cardiol 2016 Apr;23(6):636-648 [FREE Full text] [doi: 10.1177/2047487315569401] [Medline: 25687109]
3. Frederix I, Vanhees L, Dendale P, Goetschalckx K. A review of telerehabilitation for cardiac patients. J Telemed Telecare 2015 Jan;21(1):45-53. [doi: 10.1177/1357633X14562732] [Medline: 25475219]
4. de Melo Ghisi GL, Grace SL, Thomas S, Evans MF, Sawula H, Oh P. Healthcare providers' awareness of the information needs of their cardiac rehabilitation patients throughout the program continuum. Patient Educ Couns 2014 Apr;95(1):143-150. [doi: 10.1016/j.pec.2013.12.020] [Medline: 24457175]
5. Peterson JC, Link AR, Jobe JB, Winston GJ, Klimasiewfski EM, Allegrante JP. Developing self-management education in coronary artery disease. Heart Lung 2014;43(2):133-139 [FREE Full text] [doi: 10.1016/j.hrtlng.2013.11.006] [Medline: 24373484]
6. Kähkönen O, Saaranen T, Kankkunen P, Lamidi M, Kyngäs H, Miettinen H. Predictors of adherence to treatment by patients with coronary heart disease after percutaneous coronary intervention. J Clin Nurs 2018 Mar;27(5-6):989-1003. [doi: 10.1111/jocn.14153] [Medline: 29098747]
7. Liu X, Shi Y, Willis K, Wu C, Johnson M. Health education for patients with acute coronary syndrome and type 2 diabetes mellitus: an umbrella review of systematic reviews and meta-analyses. BMJ Open 2017 Oct 16;7(10):e016857 [FREE Full text] [doi: 10.1136/bmjopen-2017-016857] [Medline: 29042383]
8. Information to support well-being and service renewal. eHealth and eSocial Strategy 2020. Finnish Government, Ministry of Social Affairs and Health. Finland: Prime Minister's Office; 2015. URL: http://urn.fi/URN:ISBN:978-952-00-3575-4 [accessed 2020-09-28]
9. Neubeck L, Hansen T, Jaarsma T, Klompstra L, Gallagher R. Delivering healthcare remotely to cardiovascular patients during COVID-19 : a rapid review of the evidence. Eur J Cardiovasc Nurs 2020 Aug;19(6):486-494 [FREE Full text] [doi: 10.1177/1474515120924530] [Medline: 32380858]
10. Veen EV, Bovendeert J, Backx F, Huisstede B. E-coaching: new future for cardiac rehabilitation? A systematic review. Patient Educ Couns 2017 Dec;100(12):2218-2230. [doi: 10.1016/j.pec.2017.04.017] [Medline: 28662874]

11. Park LG, Beatty A, Stafford Z, Whooley MA. Mobile phone interventions for the secondary prevention of cardiovascular disease. Prog Cardiovasc Dis 2016;58(6):639-650. [doi: 10.1016/j.pcad.2016.03.002] [Medline: 27001245]
12. Zhang H, Jiang Y, Nguyen HD, Poo DC, Wang W. The effect of a smartphone-based coronary heart disease prevention (SBCHDP) programme on awareness and knowledge of CHD, stress, and cardiac-related lifestyle behaviours among the working population in Singapore: a pilot randomised controlled trial. Health Qual Life Outcomes 2017 Mar 14;15(1):49 [FREE Full text] [doi: 10.1186/s12955-017-0623-y] [Medline: 28288636]
13. Vosbergen S, Mulder-Wiggers JM, Lacroix JP, Kemps HM, Kraaijenhagen RA, Jaspers MW, et al. Using personas to tailor educational messages to the preferences of coronary heart disease patients. J Biomed Inform 2015 Feb;53:100-112 [FREE Full text] [doi: 10.1016/j.jbi.2014.09.004] [Medline: 25239261]
14. Blair J, Volpe M, Aggarwal B. Challenges, needs, and experiences of recently hospitalized cardiac patients and their informal caregivers. J Cardiovasc Nurs 2014;29(1):29-37 [FREE Full text] [doi: 10.1097/JCN.0b013e3182784123] [Medline: 23416934]
15. Ayasrah SM, Ahmad MM. Educational video intervention effects on periprocedural anxiety levels among cardiac catheterization patients: a randomized clinical trial. Res Theory Nurs Pract 2016;30(1):70-84. [doi: 10.1891/1541-6577.30.1.70] [Medline: 27025001]
16. Svavarsdóttir MH, Sigurðardóttir Á, Steinsbekk A. How to become an expert educator: a qualitative study on the view of health professionals with experience in patient education. BMC Med Educ 2015 May 13;15:87 [FREE Full text] [doi: 10.1186/s12909-015-0370-x] [Medline: 25966846]
17. Svavarsdóttir MH, Sigurdardottir AK, Steinsbekk A. What is a good educator? A qualitative study on the perspective of individuals with coronary heart disease. Eur J Cardiovasc Nurs 2016 Dec;15(7):513-521. [doi: 10.1177/1474515115618569] [Medline: 26588939]
18. Fors A, Gyllensten H, Swedberg K, Ekman I. Effectiveness of person-centred care after acute coronary syndrome in relation to educational level: subgroup analysis of a two-armed randomised controlled trial. Int J Cardiol 2016 Oct 15;221:957-962 [FREE Full text] [doi: 10.1016/j.ijcard.2016.07.060] [Medline: 27441475]
19. Song T, Ryu S, Lee S. U-health service for managing chronic disease: a case study on managing metabolic syndrome in a health center in South Korea. Healthc Inform Res 2011 Dec;17(4):260-266 [FREE Full text] [doi: 10.4258/hir.2011.17.4.260] [Medline: 22259728]
20. Ashrafzadeh S, Hamdy O. Patient-driven diabetes care of the future in the technology era. Cell Metab 2019 Mar 05;29(3):564-575 [FREE Full text] [doi: 10.1016/j.cmet.2018.09.005] [Medline: 30269984]
21. Honarvar B, Salehi F, Shaygani F, Hajebrahimi M, Homayounfar R, Dehghan S, et al. Opportunities and threats of electronic health in management of diabetes mellitus: an umbrella review of systematic review and meta-analysis studies. Shiraz E-Med J 2018 Nov 11. [doi: 10.5812/semj.81794]
22. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan 15;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
23. Faruque LI, Wiebe N, Ehteshami-Afshar A, Liu Y, Dianati-Maleki N, Hemmelgarn BR, Alberta Kidney Disease Network. Effect of telemedicine on glycated hemoglobin in diabetes: a systematic review and meta-analysis of randomized trials. Can Med Asso J 2017 Mar 06;189(9):341-364 [FREE Full text] [doi: 10.1503/cmaj.150885] [Medline: 27799615]
24. Heitkemper EM, Mamykina L, Travers J, Smaldone A. Do health information technology self-management interventions improve glycemic control in medically underserved adults with diabetes? A systematic review and meta-analysis. J Am Med Inform Assoc 2017 Sep 01;24(5):1024-1035. [doi: 10.1093/jamia/ocx025] [Medline: 28379397]
25. Kitsiou S, Paré G, Jaana M, Gerber B. Effectiveness of mHealth interventions for patients with diabetes: an overview of systematic reviews. PLoS One 2017;12(3):e0173160 [FREE Full text] [doi: 10.1371/journal.pone.0173160] [Medline: 28249025]
26. Hadjiconstantinou M, Byrne J, Bodicoat DH, Robertson N, Eborall H, Khunti K, et al. Do web-based interventions improve well-being in type 2 diabetes? A systematic review and meta-analysis. J Med Internet Res 2016 Oct 21;18(10):e270 [FREE Full text] [doi: 10.2196/jmir.5991] [Medline: 27769955]
27. Halldorsdottir H, Thoroddsen A, Ingadottir B. Impact of technology-based patient education on modifiable cardiovascular risk factors of people with coronary heart disease: a systematic review. Patient Educ Couns 2020 Oct;103(10):2018-2028. [doi: 10.1016/j.pec.2020.05.027] [Medline: 32595027]
28. Li R, Liang N, Bu F, Hesketh T. The effectiveness of self-management of hypertension in adults using mobile health: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Mar 27;8(3):e17776 [FREE Full text] [doi: 10.2196/17776] [Medline: 32217503]
29. Stogios N, Kaur B, Huszti E, Vasanthan J, Nolan R. Advancing digital health interventions as a clinically applied science for blood pressure reduction: a systematic review and meta-analysis. Can J Cardiol 2020 May;36(5):764-774 [FREE Full text] [doi: 10.1016/j.cjca.2019.11.010] [Medline: 32249065]
30. Sua Y, Jiang Y, Thompson D, Wang W. Effectiveness of mobile phone-based self-management interventions for medication adherence and change in blood pressure in patients with coronary heart disease: a systematic review and meta-analysis. Eur J Cardiovasc Nurs 2020 Mar;19(3):192-200. [doi: 10.1177/1474515119895678] [Medline: 31856596]

31. Timpel P, Oswald S, Schwarz P, Harst L. Mapping the evidence on the effectiveness of telemedicine interventions in diabetes, dyslipidemia, and hypertension: an umbrella review of systematic reviews and meta-analyses. J Med Internet Res 2020 Mar 18;22(3):e16791 [FREE Full text] [doi: 10.2196/16791] [Medline: 32186516]
32. Broers E, Kop W, Denollet J, Widdershoven J, Wetzels M, Ayoola I, et al. A personalized eHealth intervention for lifestyle changes in patients with cardiovascular disease: randomized controlled trial. J Med Internet Res 2020 May 22;22(5):e14570 [FREE Full text] [doi: 10.2196/14570] [Medline: 32441658]
33. Han H, Gleason K, Sun C, Miller H, Kang S, Chow S, et al. Using patient portals to improve patient outcomes: systematic review. JMIR Hum Factors 2019 Dec 19;6(4):e15038 [FREE Full text] [doi: 10.2196/15038] [Medline: 31855187]
34. Moradian S, Voelker N, Brown C, Liu G, Howell D. Effectiveness of Internet-based interventions in managing chemotherapy-related symptoms in patients with cancer: a systematic literature review. Support Care Cancer 2018 Feb;26(2):361-374. [doi: 10.1007/s00520-017-3900-8] [Medline: 28948360]
35. Aminuddin HB, Jiao N, Jiang Y, Hong J, Wang W. Effectiveness of smartphone-based self-management interventions on self-efficacy, self-care activities, health-related quality of life and clinical outcomes in patients with type 2 diabetes: a systematic review and meta-analysis. Int J Nurs Stud 2019 Feb 08. [doi: 10.1016/j.ijnurstu.2019.02.003] [Medline: 30827741]
36. Jin K, Khonsari S, Gallagher R, Gallagher P, Clark AM, Freedman B, et al. Telehealth interventions for the secondary prevention of coronary heart disease: a systematic review and meta-analysis. Eur J Cardiovasc Nurs 2019 Apr;18(4):260-271. [doi: 10.1177/1474515119826510] [Medline: 30667278]
37. Ma Y, Cheng HY, Cheng L, Sit JW. The effectiveness of electronic health interventions on blood pressure control, self-care behavioural outcomes and psychosocial well-being in patients with hypertension: a systematic review and meta-analysis. Int J Nurs Stud 2019 Apr;92:27-46. [doi: 10.1016/j.ijnurstu.2018.11.007] [Medline: 30690164]
38. Triberti S, Savioni L, Sebri V, Pravettoni G. eHealth for improving quality of life in breast cancer patients: a systematic review. Cancer Treat Rev 2019 Mar;74:1-14. [doi: 10.1016/j.ctrv.2019.01.003] [Medline: 30658289]
39. Woolley AK, Hadjiconstantinou M, Davies M, Khunti K, Seidu S. Online patient education interventions in type 2 diabetes or cardiovascular disease: a systematic review of systematic reviews. Prim Care Diabetes 2019 Feb;13(1):16-27. [doi: 10.1016/j.pcd.2018.07.011] [Medline: 30131299]
40. Lunde P, Nilsson BB, Bergland A, Kværner KJ, Bye A. The effectiveness of smartphone apps for lifestyle improvement in noncommunicable diseases: systematic review and meta-analyses. J Med Internet Res 2018 May 04;20(5):e162 [FREE Full text] [doi: 10.2196/jmir.9751] [Medline: 29728346]
41. Xiong S, Berkhouse H, Schooler M, Pu W, Sun A, Gong E, et al. Effectiveness of mHealth interventions in improving medication adherence among people with hypertension: a systematic review. Curr Hypertens Rep 2018 Aug 07;20(10):86. [doi: 10.1007/s11906-018-0886-7] [Medline: 30088110]
42. Gandhi S, Chen S, Hong L, Sun K, Gong E, Li C, et al. Effect of mobile health interventions on the secondary prevention of cardiovascular disease: systematic review and meta-analysis. Can J Cardiol 2017 Feb;33(2):219-231. [doi: 10.1016/j.cjca.2016.08.017] [Medline: 27956043]
43. Hanlon P, Daines L, Campbell C, McKinstry B, Weller D, Pinnock H. Telehealth interventions to support self-management of long-term conditions: a systematic metareview of diabetes, heart failure, asthma, chronic obstructive pulmonary disease, and cancer. J Med Internet Res 2017 May 17;19(5):e172 [FREE Full text] [doi: 10.2196/jmir.6688] [Medline: 28526671]
44. Rush KL, Hatt L, Janke R, Burton L, Ferrier M, Tetrault M. The efficacy of telehealth delivered educational approaches for patients with chronic diseases: a systematic review. Patient Educ Couns 2018 Aug;101(8):1310-1321. [doi: 10.1016/j.pec.2018.02.006] [Medline: 29486994]
45. Beishuizen CR, Stephan BC, Van Gool WA, Brayne C, Peters RJ, Andrieu S, et al. Web-based interventions targeting cardiovascular risk factors in middle-aged and older people: a systematic review and meta-analysis. J Med Internet Res 2016 Mar 11;18(3):e55 [FREE Full text] [doi: 10.2196/jmir.5218] [Medline: 26968879]
46. Rawstorn JC, Gant N, Direito A, Beckmann C, Maddison R. Telehealth exercise-based cardiac rehabilitation: a systematic review and meta-analysis. Heart 2016 Dec 01;102(15):1183-1192. [doi: 10.1136/heartjnl-2015-308966] [Medline: 26936337]
47. Devi R, Singh SJ, Powell J, Fulton EA, Igbinedion E, Rees K. Internet-based interventions for the secondary prevention of coronary heart disease. Cochrane Database Syst Rev 2015(12):CD009386. [doi: 10.1002/14651858.CD009386.pub2] [Medline: 26691216]
48. Timpel P, Harst L. Research implications for future telemedicine studies and innovations in diabetes and hypertension-a mixed methods study. Nutrients 2020 May 08;12(5):1340 [FREE Full text] [doi: 10.3390/nu12051340] [Medline: 32397096]
49. Systematic reviews. CRD's guidance for undertaking reviews in health care. Centre for Reviews and Dissemination. 2009. URL: http://www.york.ac.uk/crd/SysRev/!SSL!/WebHelp/SysRev3.htm [accessed 2020-09-29]
50. Aromataris E, Munn Z. JBI Manual for Evidence Synthesis. Australia: JBI - The University of Adelaide; 2020.
51. Moher D, Liberati A, Tetzlaff J, Altman D, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
52. Checklist for randomized controlled trials: critical appraisal tools for use in JBI systematic reviews. Joanna Briggs Institute. 2020. URL: https://joannabriggs.org/sites/default/files/2020-08/Checklist_for_RCTs.pdf [accessed 2020-09-29]

53. Børøsund E, Cvancarova M, Moore SM, Ekstedt M, Ruland CM. Comparing effects in regular practice of e-communication and web-based self-management support among breast cancer patients: preliminary results from a randomized controlled trial. J Med Internet Res 2014;16(12):e295 [FREE Full text] [doi: 10.2196/jmir.3348] [Medline: 25525672]
54. Choi BG, Dhawan T, Metzger K, Marshall L, Akbar A, Jain T, et al. Image-based mobile system for dietary management in an American cardiology population: pilot randomized controlled trial to assess the efficacy of dietary coaching delivered via a smartphone app versus traditional counseling. JMIR Mhealth Uhealth 2019 Apr 23;7(4):e10755 [FREE Full text] [doi: 10.2196/10755] [Medline: 31012860]
55. Elkjaer M, Shuhaibar M, Burisch J, Bailey Y, Scherfig H, Laugesen B, et al. E-health empowers patients with ulcerative colitis: a randomised controlled trial of the web-guided 'Constant-care' approach. Gut 2010 Dec;59(12):1652-1661. [doi: 10.1136/gut.2010.220160] [Medline: 21071584]
56. Habibović M, Denollet J, Cuijpers P, van der Voort PH, Herrman J, Bouwels L, et al. Web-based distress management for implantable cardioverter defibrillator patients: a randomized controlled trial. Health Psychol 2017 Apr;36(4):392-401. [doi: 10.1037/hea0000451] [Medline: 28192003]
57. Keyserling TC, Sheridan SL, Draeger LB, Finkelstein EA, Gizlice Z, Kruger E, et al. A comparison of live counseling with a web-based lifestyle and medication intervention to reduce coronary heart disease risk: a randomized clinical trial. JAMA Intern Med 2014 Jul;174(7):1144-1157 [FREE Full text] [doi: 10.1001/jamainternmed.2014.1984] [Medline: 24861959]
58. Kim JY, Wineinger NE, Steinhubl SR. The influence of wireless self-monitoring program on the relationship between patient activation and health behaviors, medication adherence, and blood pressure levels in hypertensive patients: a substudy of a randomized controlled trial. J Med Internet Res 2016 Jun 22;18(6):e116 [FREE Full text] [doi: 10.2196/jmir.5429] [Medline: 27334418]
59. Lambert TE, Harvey LA, Avdalis C, Chen LW, Jeyalingam S, Pratt CA, et al. An app with remote support achieves better adherence to home exercise programs than paper handouts in people with musculoskeletal conditions: a randomised trial. J Physiother 2017 Jul;63(3):161-167 [FREE Full text] [doi: 10.1016/j.jphys.2017.05.015] [Medline: 28662834]
60. Mata J, Pecorelli N, Kaneva P, Moldoveanu D, Gosselin-Tardiff A, Alhashemi M, et al. A mobile device application (app) to improve adherence to an enhanced recovery program for colorectal surgery: a randomized controlled trial. Surg Endosc 2019 May 13:742-751. [doi: 10.1007/s00464-019-06823-w] [Medline: 31087175]
61. Sharara AI, Chalhoub JM, Beydoun M, Shayto RH, Chehab H, Harb AH, et al. A customized mobile application in colonoscopy preparation: a randomized controlled trial. Clin Transl Gastroenterol 2017 Jan 05;8(1):e211 [FREE Full text] [doi: 10.1038/ctg.2016.65] [Medline: 28055031]
62. Strøm J, Nielsen CV, Jørgensen LB, Andersen NT, Laursen M. A web-based platform to accommodate symptoms of anxiety and depression by featuring social interaction and animated information in patients undergoing lumbar spine fusion: a randomized clinical trial. Spine J 2019 May;19(5):827-839 [FREE Full text] [doi: 10.1016/j.spinee.2018.11.011] [Medline: 30500464]
63. White V, Farrelly A, Pitcher M, Hill D. Does access to an information-based, breast cancer specific website help to reduce distress in young women with breast cancer? Results from a randomised trial. Eur J Cancer Care (Engl) 2018 Nov;27(6):e12897. [doi: 10.1111/ecc.12897] [Medline: 30137657]
64. Widmer RJ, Allison TG, Lennon R, Lopez-Jimenez F, Lerman LO, Lerman A. Digital health intervention during cardiac rehabilitation: a randomized controlled trial. Am Heart J 2017 Jun;188:65-72. [doi: 10.1016/j.ahj.2017.02.016] [Medline: 28577682]
65. Yun YH, Lee KS, Kim Y, Park SY, Lee ES, Noh D, et al. Web-based tailored education program for disease-free cancer survivors with cancer-related fatigue: a randomized controlled trial. J Clin Oncol 2012 Apr 20;30(12):1296-1303 [FREE Full text] [doi: 10.1200/JCO.2011.37.2979] [Medline: 22412149]
66. Urech C, Grossert A, Alder J, Scherer S, Handschin B, Kasenda B, et al. Web-based stress management for newly diagnosed patients with cancer (STREAM): a randomized, wait-list controlled intervention study. J Clin Oncol 2018 Mar 10;36(8):780-788 [FREE Full text] [doi: 10.1200/JCO.2017.74.8491] [Medline: 29369731]
67. Persell SD, Peprah YA, Lipiszko D, Lee JY, Li JJ, Ciolino JD, et al. Effect of home blood pressure monitoring via a smartphone hypertension coaching application or tracking application on adults with uncontrolled hypertension: a randomized clinical trial. JAMA Netw Open 2020 Mar 02;3(3):e200255 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.0255] [Medline: 32119093]
68. Liu Y, Chen J, Lamb KV, Wu P, Chang P, Cui Y, et al. Smartphone-based self-empowerment app on secondary prevention of patients with cardiovascular disease. Stud Health Technol Inform 2019 Aug 21;264:1712-1713. [doi: 10.3233/SHTI190610] [Medline: 31438306]
69. Greer JA, Jacobs J, Pensak N, MacDonald JJ, Fuh C, Perez GK, et al. Randomized trial of a tailored cognitive-behavioral therapy mobile application for anxiety in patients with incurable cancer. Oncologist 2019 Aug;24(8):1111-1120. [doi: 10.1634/theoncologist.2018-0536] [Medline: 30683710]
70. Handa S, Okuyama H, Yamamoto H, Nakamura S, Kato Y. Effectiveness of a smartphone application as a support tool for patients undergoing breast cancer chemotherapy: a randomized controlled trial. Clin Breast Cancer 2020 Jun;20(3):201-208. [doi: 10.1016/j.clbc.2020.01.004] [Medline: 32201165]

71. Yu C, Liu C, Du J, Liu H, Zhang H, Zhao Y, MISSION-2 Collaborative Group. Smartphone-based application to improve medication adherence in patients after surgical coronary revascularization. Am Heart J 2020 Oct;228:17-26. [doi: 10.1016/j.ahj.2020.06.019] [Medline: 32745732]
72. Liu S, Tanaka R, Barr S, Nolan RP. Effects of self-guided e-counseling on health behaviors and blood pressure: results of a randomized trial. Patient Educ Couns 2020 Mar;103(3):635-641. [doi: 10.1016/j.pec.2019.10.007] [Medline: 31669047]
73. Foley NM, O'Connell EP, Lehane EA, Livingstone V, Maher B, Kaimkhani S, et al. PATI: Patient accessed tailored information: a pilot study to evaluate the effect on preoperative breast cancer patients of information delivered via a mobile application. Breast 2016 Dec;30:54-58. [doi: 10.1016/j.breast.2016.08.012] [Medline: 27611236]
74. Guo Y, Chen Y, Lane DA, Liu L, Wang Y, Lip GY. Mobile health technology for atrial fibrillation management integrating decision support, education, and patient involvement: MAF app trial. Am J Med 2017 Dec;130(12):1388-1396 [FREE Full text] [doi: 10.1016/j.amjmed.2017.07.003] [Medline: 28847546]
75. Helzlsouer K, Appling S, Gallicchio L, Henninger D, MacDonald R, Manocheh S, et al. A pilot study of a virtual navigation program to improve treatment adherence among low-income breast cancer patients. J Oncol Navigat Surv 2016;7(7):20-29 [FREE Full text]
76. Petzel SV, Vogel RI, Cragg J, McClellan M, Chan D, Jacko JA, et al. Effects of web-based instruction and patient preferences on patient-reported outcomes and learning for women with advanced ovarian cancer: a randomized controlled trial. J Psychosoc Oncol 2018;36(4):503-519 [FREE Full text] [doi: 10.1080/07347332.2018.1457125] [Medline: 29791275]
77. Fang S, Wang Y, Lu W, Lee K, Kuo Y, Fetzer SJ. Long-term effectiveness of an E-based survivorship care plan for breast cancer survivors: a quasi-experimental study. Patient Educ Couns 2020 Mar;103(3):549-555. [doi: 10.1016/j.pec.2019.09.012] [Medline: 31558323]
78. Park OL, Kim SR. Integrated self-management program effects on hemodialysis patients: a quasi-experimental study. Jpn J Nurs Sci 2019 Oct;16(4):396-406. [doi: 10.1111/jjns.12249] [Medline: 30669185]
79. Hoffmann T, Glasziou P, Boutron I, Milne R, Perera R, Moher D, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. Br Med J 2014 Mar 07;348:g1687 [FREE Full text] [doi: 10.1136/bmj.g1687] [Medline: 24609605]
80. Xu L, Li F, Zhou C, Li J, Hong C, Tong Q. The effect of mobile applications for improving adherence in cardiac rehabilitation: a systematic review and meta-analysis. BMC Cardiovasc Disord 2019 Jul 12;19(1):166 [FREE Full text] [doi: 10.1186/s12872-019-1149-5] [Medline: 31299903]
81. Fredericks S, Martorella G, Catallo C. A systematic review of web-based educational interventions. Clin Nurs Res 2015 Feb;24(1):91-113. [doi: 10.1177/1054773814522829] [Medline: 24571963]
82. Öberg U, Isaksson U, Jutterström L, Orre CJ, Hörnsten A. Perceptions of persons with type 2 diabetes treated in swedish primary health care: qualitative study on using eHealth services for self-management support. JMIR Diabetes 2018 Mar 12;3(1):e7 [FREE Full text] [doi: 10.2196/diabetes.9059] [Medline: 30291075]
83. Richards R, Kinnersley P, Brain K, McCutchan G, Staffurth J, Wood F. Use of mobile devices to help cancer patients meet their information needs in non-inpatient settings: systematic review. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e10026 [FREE Full text] [doi: 10.2196/10026] [Medline: 30552082]
84. Pouls B, Vriezekolk J, Bekker C, Linn A, van Onzenoort HA, Vervloet M, et al. Effect of interactive eHealth interventions on improving medication adherence in adults with long-term medication: systematic review. J Med Internet Res 2021 Jan 08;23(1):e18901 [FREE Full text] [doi: 10.2196/18901] [Medline: 33416501]
85. Brørs G, Pettersen T, Hansen T, Fridlund B, Hølvold LB, Lund H, et al. Modes of e-Health delivery in secondary prevention programmes for patients with coronary artery disease: a systematic review. BMC Health Serv Res 2019 Jun 10;19(1):364 [FREE Full text] [doi: 10.1186/s12913-019-4106-1] [Medline: 31182100]
86. Penedo FJ, Oswald LB, Kronenfeld JP, Garcia SF, Cella D, Yanez B. The increasing value of eHealth in the delivery of patient-centred cancer care. Lancet Oncol 2020 May;21(5):240-251. [doi: 10.1016/s1470-2045(20)30021-8]
87. Bruce C, Harrison P, Giammattei C, Desai S, Sol J, Jones S, et al. Evaluating patient-centered mobile health technologies: definitions, methodologies, and outcomes. JMIR Mhealth Uhealth 2020 Nov 11;8(11):e17577 [FREE Full text] [doi: 10.2196/17577] [Medline: 33174846]

Abbreviations

HADS: Hospital Anxiety and Depression Scale
mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by R Kukafka; submitted 30.04.21; peer-reviewed by S Fredericks, X Guo; comments to author 28.06.21; accepted 21.11.21; published 06.01.22.

Please cite as:
Paalimäki-Paakki K, Virtanen M, Henner A, Nieminen MT, Kääriäinen M
Effectiveness of Digital Counseling Environments on Anxiety, Depression, and Adherence to Treatment Among Patients Who Are Chronically Ill: Systematic Review
J Med Internet Res 2022;24(1):e30077
URL: https://www.jmir.org/2022/1/e30077
doi:10.2196/30077
PMID:34989681

Review

Digital Behavior Change Interventions for the Prevention and Management of Type 2 Diabetes: Systematic Market Analysis

Roman Keller[1,2], MSc; Sven Hartmann[3], MA; Gisbert Wilhelm Teepe[4], MSc; Kim-Morgaine Lohse[4], MSc; Aishah Alattas[1], MSc; Lorainne Tudor Car[5,6], PhD, MD; Falk Müller-Riemenschneider[2,7], MD, PhD; Florian von Wangenheim[1,4], PhD; Jacqueline Louise Mair[1,2*], PhD; Tobias Kowatsch[1,3,4*], PhD

[1]Future Health Technologies Programme, Campus for Research Excellence and Technological Enterprise, Singapore-ETH Centre, Singapore, Singapore
[2]Saw Swee Hock School of Public Health, National University of Singapore, Singapore, Singapore
[3]Centre for Digital Health Interventions, Institute of Technology Management, University of St Gallen, St Gallen, Switzerland
[4]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland
[5]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore
[6]Department of Primary Care and Public Health, School of Public Health, Imperial College London, London, United Kingdom
[7]Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[*]these authors contributed equally

Corresponding Author:
Roman Keller, MSc
Future Health Technologies Programme
Campus for Research Excellence and Technological Enterprise
Singapore-ETH Centre
1 Create Way
CREATE Tower, #06-01
Singapore, 138602
Singapore
Phone: 65 82645302
Email: roman.keller@sec.ethz.ch

Abstract

Background: Advancements in technology offer new opportunities for the prevention and management of type 2 diabetes. Venture capital companies have been investing in digital diabetes companies that offer digital behavior change interventions (DBCIs). However, little is known about the scientific evidence underpinning such interventions or the degree to which these interventions leverage novel technology-driven automated developments such as conversational agents (CAs) or just-in-time adaptive intervention (JITAI) approaches.

Objective: Our objectives were to identify the top-funded companies offering DBCIs for type 2 diabetes management and prevention, review the level of scientific evidence underpinning the DBCIs, identify which DBCIs are recognized as evidence-based programs by quality assurance authorities, and examine the degree to which these DBCIs include novel automated approaches such as CAs and JITAI mechanisms.

Methods: A systematic search was conducted using 2 venture capital databases (Crunchbase Pro and Pitchbook) to identify the top-funded companies offering interventions for type 2 diabetes prevention and management. Scientific publications relating to the identified DBCIs were identified via PubMed, Google Scholar, and the DBCIs' websites, and data regarding intervention effectiveness were extracted. The Diabetes Prevention Recognition Program (DPRP) of the Center for Disease Control and Prevention in the United States was used to identify the recognition status. The DBCIs' publications, websites, and mobile apps were reviewed with regard to the intervention characteristics.

Results: The 16 top-funded companies offering DBCIs for type 2 diabetes received a total funding of US $2.4 billion as of June 15, 2021. Only 4 out of the 50 identified publications associated with these DBCIs were fully powered randomized controlled trials (RCTs). Further, 1 of those 4 RCTs showed a significant difference in glycated hemoglobin A_{1c} (HbA_{1c}) outcomes between the intervention and control groups. However, all the studies reported HbA_{1c} improvements ranging from 0.2% to 1.9% over the course of 12 months. In addition, 6 interventions were fully recognized by the DPRP to deliver evidence-based programs, and 2 interventions had a pending recognition status. Health professionals were included in the majority of DBCIs (13/16, 81%,),

whereas only 10% (1/10) of accessible apps involved a CA as part of the intervention delivery. Self-reports represented most of the data sources (74/119, 62%) that could be used to tailor JITAIs.

Conclusions: Our findings suggest that the level of funding received by companies offering DBCIs for type 2 diabetes prevention and management does not coincide with the level of evidence on the intervention effectiveness. There is considerable variation in the level of evidence underpinning the different DBCIs and an overall need for more rigorous effectiveness trials and transparent reporting by quality assurance authorities. Currently, very few DBCIs use automated approaches such as CAs and JITAIs, limiting the scalability and reach of these solutions.

(J Med Internet Res 2022;24(1):e33348) doi:10.2196/33348

KEYWORDS

digital health companies; health care; type 2 diabetes; prevention; management; conversational agent; digital behavior change intervention; investment; just-in-time adaptive intervention; digital health; diabetes; agent; behavior

Introduction

In 2019, approximately 463 million adults were estimated to be living with diabetes [1]. This estimate is expected to rise to more than 700 million by 2045 [1]. More than 90% of this burden is caused by type 2 diabetes. [1,2]. Over 1 million deaths worldwide were attributed to this condition in 2017 alone, making it the ninth leading cause of mortality [3]. Diabetes is also a leading source of global health expenditure with an estimated annual cost of US $760 billion in high- and low-income countries, including the United States (US $259 billion), China (US $109 billion), and Brazil (US $52 billion) [1].

Guidelines for the prevention and management of type 2 diabetes include specific recommendations for lifestyle behavior changes such as diet, exercise, smoking cessation, and the addition of oral antidiabetic agents or insulin therapy in some cases [4,5]. Traditionally, diabetes prevention and self-management education programs have been delivered in person with individual or face-to-face group interactions between health professionals and participants [6]. However, traditional in-person approaches have been hampered by low uptake and engagement rates [7]. Qualitative literature suggests that participants often find face-to-face programs difficult to attend because of issues with the timing of the courses, lack of transport, family and work commitments, or negative feelings toward participating in groups [8]. More recently, digital behavior change interventions (DBCIs) for diabetes prevention and management have emerged as potentially effective, scalable, and low-cost options to provide behavioral counseling when in-person programs are not accessible or attractive [9-11].

DBCIs are interventions that use digital technology to encourage and support behavior change that will maintain or improve health through the prevention and management of health problems and can, for example, be delivered through computer programs, websites, mobile apps, or wearable devices [12]. DBCIs may involve telehealth elements such as remote monitoring by health professionals who provide virtual support, either individually or in groups, or fully automated interventions that are based on algorithms [13]. DBCIs are becoming increasingly automated, interactive, and personalized because they use self-reports of users or sensor data to tailor feedback without the need for inputs from health professionals [14]. This development is facilitated by new technology-driven developments such as conversational agents (CAs) and just-in-time adaptive interventions (JITAIs). CAs, also known as chatbots, are computer systems that imitate human conversation using text or spoken language and can offer personalized human-like interactions [15-18]. Evidence from interventions using CAs show promising findings in terms of patient satisfaction [19], treatment success [20], and the capability to build work alliances with the patient [21-23]. CAs can also foster experiences equivalent to those offered by human coaches but with the additional advantage of being persistent and more consistent in providing choices that cultivate user autonomy [24]. This makes the use of CAs in DBCIs an encouraging component to complement or replace the need for human health professionals in intervention delivery.

Moreover, recent advances in wireless devices and mobile technology have enabled the design of JITAIs that can provide behavior change support at opportune moments and in response to an individual's changing contexts [25-27]. More specifically, JITAIs adapt the provision of intervention content (eg, the type, timing, and intensity) by measuring the health condition or patient behavior with mobile technology such as smartphones, sensors, and software analytics to deliver intervention content at the time and in the context that the person needs it the most, and this is likely to improve health-related behaviors [25,27-29].

Novel technology-driven opportunities for DBCIs in diabetes care have attracted various health care stakeholders such as investors, health insurance companies, researchers, physicians, and patients [30]. The global market for digital diabetes care is rapidly growing and is expected to be worth US $1.5 billion in 2024 [31]. In 2018 alone, venture capital companies invested a record US $417 million into digital diabetes companies, a 12-fold increase in funding compared to 2013 [32]. However, little is known about the DBCIs provided by companies that have a substantial impact on the market, including the content of the interventions, how effective they are in managing and preventing type 2 diabetes, and the degree to which these interventions leverage new technology-driven developments such as CAs or JITAIs.

The aim of this paper is to systematically review the solutions provided by the top-funded companies offering DBCIs for type 2 diabetes prevention and management with a particular focus on how new technological developments, such as CAs and JITAIs, are being used to automate and scale-up intervention

delivery. Therefore, the paper has the following objectives: (1) to identify the top-funded companies offering DBCIs for type 2 diabetes management and prevention, (2) to appraise the level of evidence to support these DBCIs in the form of peer-reviewed publications and recognition by national authorities for delivering evidence-based programs, and (3) to describe the characteristics of these DBCIs, with particular focus on the use of automation involved in the DBCIs by investigating the use of CAs, involvement of human health professionals, and what as well as how health and behavioral outcomes are measured that could be used to tailor JITAIs.

Methods

Searches

Companies

Digital health companies offering DBCIs were identified using 2 venture capital databases, Crunchbase Pro and Pitchbook [33,34]. Both databases are among the most comprehensive and accurate venture capital databases and are commonly used as data sources for academic reports and by investors [35]. We define digital health companies as companies that build and sell digital health products or services according to the definition of Safavi et al [36].

Searches were carried out on July 23, 2020, and they were updated on April 8, 2021 (Crunchbase Pro only). The total funding amount was last updated on June 15, 2021 using Crunchbase Pro). In case of conflicting funding information between the 2 databases, Crunchbase Pro data were reported, as Crunchbase Pro has better coverage than Pitchbook with respect to the financing rounds and total capital committed [35]. The search strategy included an extensive list of terms describing the constructs “verticals, methods, and industries,” “diabetes,” and “management and prevention.” The overview of the complete search strategy used for Crunchbase and Pitchbook is given in Table 1.

Table 1. Search strategy used in Crunchbase Pro and Pitchbook.

Search category	Search terms
1. Verticals, methods, and industries	Monitoring Equipment OR diagnostic OR HealthTech OR healthcare devices OR connected health* OR Therapeutic Devices OR Digital Health OR digital health* OR health* technology OR health* app* OR wearables OR Mobile health OR mhealth OR mobile app OR personal health OR virtual care OR e-health OR assistive technology OR telehealth OR telemedicine OR health* platform OR healthcare it OR data management OR Artificial Intelligence & Machine Learning OR Cloud data services OR analytics OR health* diagnostics OR Big Data OR information OR digital OR data OR biometrics OR home health care OR medtech OR self-monitoring
2. Diabetes	obesity OR blood sugar OR blood glucose OR insulin OR diabet*
3. Management and prevention	diabetes management OR diabetes treatment OR diabetes control OR diabetes monitoring OR blood sugar monitoring OR disease monitoring OR disease management OR risk reduction OR disease prevention OR diabetes prevention OR prevention OR prediabet*

Inclusion and Exclusion Criteria

We were interested in the companies having a substantial impact on the market and their ability to develop evidence-based solutions. Therefore, we decided to limit the scope of the analysis to the 15 top-funded companies defined as the leading companies in terms of the total funding amount, given that these companies are likely best equipped to develop and evaluate their interventions.

Companies were included if they (1) offered a DBCI for the prevention or management of type 2 diabetes and (2) involved a mobile app as the main intervention component. Companies were excluded if their DBCI (1) did not predominantly involve behavior change, (2) did not involve a mobile app as the main intervention component, and (3) did not focus on type 2 diabetes. We also excluded companies where the targeted conditions of the companies’ DBCIs were not clearly identifiable.

Company Selection

Following the removal of duplicates, companies were ranked in the order of their funding amount. Company screening was conducted by screening from the most to the least funded companies until 15 companies eligible for inclusion in the study were identified. All the remaining companies were excluded due to insufficient funding amount. The list of the identified companies was reviewed by 3 experts with extensive industry and academic experience in the fields of digital health and type 2 diabetes to confirm that all relevant companies, covering the current market, had been identified through database searching. The experts included 2 scientific researchers with over 10 years of work experience with DBCIs at universities in the United Kingdom and United States and 1 industry expert with several years of work experience at one of the global market leaders for diabetes management systems in Germany.

Publications

We searched PubMed and Google Scholar for scientific articles published up to April 30, 2021, using search terms “Name_Intervention” AND (Smartphone OR Application OR App OR Intervention OR Mobile Health) relating to the identified company’s DBCI. In addition, we identified studies by screening the websites of the companies for publication references.

Inclusion and Exclusion Criteria

To investigate the impact of the included DBCIs on health or behavioral outcomes in the study population, we included publications reporting quantitative results of experimental trials. Therefore, we excluded studies that did not involve effectiveness outcomes and those that did not report quantitative results.

Furthermore, we excluded protocol studies and studies that targeted conditions other than type 2 diabetes.

DBCIs

All the identified DBCIs included a mobile app as the main form of intervention delivery. We searched and downloaded all the identified apps from the 2 most popular app stores, Google Play Store and Apple App Store [37], between October 12, 2020, and April 10, 2021. If an app was not accessible, the companies were approached via email to request access. If no reply was received for the first email, a follow-up email was sent 2 weeks later. We also reviewed the DBCIs' and companies' websites as well as the identified publications for information on the characteristics of the DBCIs. Additional hardware devices such as activity trackers, blood glucose meters, wireless scales, or blood pressure devices that came as a part of the intervention program were not available and were therefore not reviewed.

Data Extraction

Data extraction of companies, publications, and DBCIs was performed by 2 independent investigators (RK and SH). Disagreements were discussed and resolved by consensus. If no agreement was possible, disagreements were resolved through discussion with a third reviewer (GWT). Data extraction was performed using the Covidence Systematic Review software (Veritas Health Innovation Ltd) [38].

Companies

The extracted data for each company included the founding year, total funding amount, number of employees, and company headquarter location.

Publications

From the identified publications, we extracted the publication year, study design, number of participants, measured outcomes, quality of evidence (using the criteria of the US Preventive Services Task Force), journal impact factor, comparison to other treatment methods, and study findings. Similar to Safavi et al [36], the quality of individual studies was defined according to the USPSTF hierarchy of research design as follows: Level 1 includes evidence obtained from properly powered and conducted randomized controlled trials (RCTs), well-conducted systematic reviews, or meta-analyses of homogeneous RCTs. Level 2 includes evidence obtained from well-designed controlled trials without randomization, well-designed cohort or case-control analysis studies, or multiple time-series designs with or without the intervention or dramatic results in uncontrolled studies of large magnitude. Level 3 includes opinions of respected authorities, based on clinical experience or descriptive studies, or reports of expert committees [39]. As we were interested in the best available scientific evidence regarding the interventions, we extracted the results of publications with quality level 1. We specifically examined the primary outcome(s) from RCTs that were powered to detect change.

DBCIs

For each DBCI, we extracted the name of the intervention, name of the app, app accessibility information, number of app downloads (from Google Play Store only, as this information is not available on the Apple App Store), operating systems, cost, addressed category of the health care continuum (management or prevention), and the involvement of health professionals. For each DBCI with app access, we also extracted information on the availability of a CA and the measured health and behavioral outcomes. We were particularly interested in what and how health and behavioral outcomes were measured and if they could potentially be used to tailor JITAIs. Health and behavioral outcomes were defined as any biomarkers or health behaviors relevant for diabetes care such as diet, physical activity, or blood glucose tracking. Measurements included self-report data or sensor and device analytics [40-42]. More information on the framework used to assess the measurements of health and behavioral outcomes can be found in Multimedia Appendix 1.

We were also interested in whether the DBCIs were recognized by a national authority as an evidence-based program. For this purpose, we used the Diabetes Prevention Recognition Program (DPRP) developed by the US Centers for Disease Control and Prevention (CDC) [43]. The DPRP is the quality assurance arm of the National Diabetes Prevention Program, which is a partnership of public and private organizations that aim to prevent or delay type 2 diabetes [43]. Through the DPRP, the US CDC recognizes organizations that have demonstrated their ability to deliver an effective lifestyle change program. The organizations are required to use a CDC-approved curriculum and can deliver the intervention either in person by employing a trained human health coach or through a virtual setting with interaction involving a lifestyle coach [44]. The organizations are evaluated regularly based on the participant data submitted to the DPRP. These data need to fulfill a set of requirements, including a reduction in the risk of diabetes by achieving improvements in participant outcomes such as weight loss or glycated hemoglobin (HbA_{1c}) reductions [44].

Data Synthesis

The information extracted from the companies, publications, and DBCIs was summarized narratively.

Results

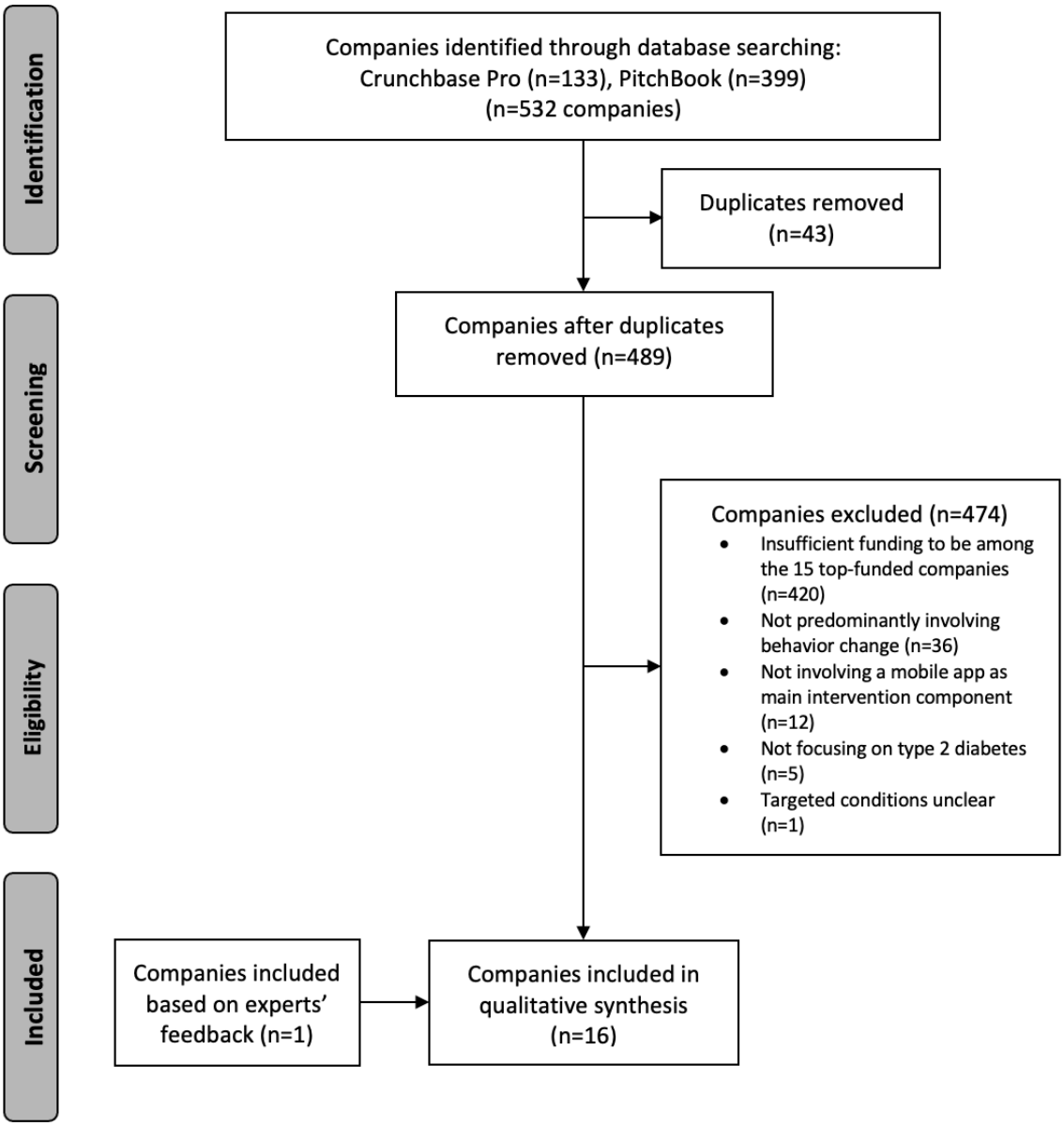
Selection and Inclusion of Companies

The search yielded a total of 133 companies on Crunchbase Pro and 399 companies on Pitchbook. After removal of duplicates, 489 companies were eligible for screening. After screening, 54 companies were found to be ineligible for study inclusion, with the most common reason being not predominantly involving behavior change (36/54, 67%). Of the remaining 435 companies, 420 were excluded due to insufficient funding to be among the 15 top-funded companies. An additional company (KKT Technology Pte Ltd) was included on the recommendation of the independent experts, ultimately resulting in 16 companies eligible for study inclusion. Figure 1 outlines the selection process and reasons for exclusion. All the DBCIs of the included companies were available in English language.

The apps of 6 DBCIs were not accessible to the study authors (Virta, Dario, Welldoc, Liva, Twin, and Sweetch) because they

were only available with a subscription service, in a specific geographic region, with an employer subscription, or when referred by a physician. Therefore, no information on the health and behavioral outcomes, measurements, or availability of CAs is provided on these apps within the results.

Figure 1. Flowchart of the company selection process.



Company Characteristics

The funding amount of the 16 top-funded digital health companies chosen for inclusion in the analysis ranged from US $657.3 to 15.5 million, totaling US $2.355 billion, as indicated in Table 2. Moreover, 11 companies (69%) were headquartered in the United States, 2 (13%) in the United Kingdom (13%), and 1 each in Denmark, Israel, and Singapore (6%). The year of founding ranged between 2005 and 2018, with 81% (13/16 companies) founded from 2011 onward. Additional information regarding the companies' characteristics can be found in Multimedia Appendix 2.

Table 2. Overview of funding amounts determined for the included companies and scientific evidence obtained for the included digital behavior change interventions.

Company name	Funding (million US$)	DBCI[a] name	Number of publications categorized by evidence level[b]			DPRP[c] recognition[d]
			Level 1	Level 2	Level 3	
Noom Inc	657.3	Noom	1	7	0	Full
Virta Health Corp	373	Virta	0	7	0	None
Omada Health Inc	256.5	Omada	0	11	0	Full
Livongo Health Inc	235	Livongo	1	3	0	Full
Vida Health Inc	188	Vida	0	2	0	Full
DarioHealth Corp	169	Dario	0	0	0	None
Informed Data Systems Inc (One Drop)	106.2	One Drop	0	2	0	Pending
Lark Technologies Inc	95.7	Lark	0	1	0	Full
Welldoc Inc	55.2	BlueStar	1	5	0	Pending
Liva Healthcare ApS	43.5	Liva	1	3	0	None
Twin Health Inc	43.5	Twin	0	1	0	None
Oviva Inc	33	Oviva	0	2	0	None
KKT Technology Pte Ltd (Holmusk)	31.3	GlycoLeap	0	1	0	None
Sweetch Health Ltd	27.5	Sweetch	0	1	0	None
Nemaura Medical Inc	25	BEATdiabetes	0	0	0	None
Fruit Street Health Inc	15.5	Fruit Street	0	0	0	Full

[a]DBCI: digital behavior change intervention.

[b]Publication evidence level determined using the criteria of the US Preventive Services Task Force.

[c]DPRP: Diabetes Prevention Recognition Program.

[d]Recognition was established as none, pending, preliminary, or full, in line with the DPRP.

Scientific Evidence

Totally 50 published studies related to the 16 companies' DBCIs focusing on effectiveness were identified, as shown in Table 2. Further details on the study characteristics are available in Multimedia Appendix 3. The publication dates ranged from 2008 to 2021, with 86% (43/50) of the studies published from 2016 onward. The sample size of each study ranged from 16 to 35,921 participants. Out of the 50 studies, only 4 (8%) had quality level 1, evaluating DBCIs Noom, Livongo, BlueStar, and Liva. The remaining 46 studies (92%) had quality level 2. No studies were found for interventions Dario, BEATdiabetes, and Fruit Street. For 8 DBCIs, the recognition status in the DPRP of the US CDC was available, of which 6 DBCIs achieved full CDC recognition (Noom, Omada, Livongo, Vida, Lark, and Fruit Street), and 2 DBCIs had a pending recognition status (One Drop and BlueStar).

Effectiveness of DBCIs

Of the 4 identified studies with quality level 1, 3 were RCTs having a duration of 12 months involving interventions Noom, BlueStar, and Liva [45-47], whereas 1 study involving Livongo [48] was a 6-month-long intervention tested within a randomized crossover trial spanning 12 months, with crossover at 6 months. BlueStar was the only intervention that resulted in a significantly greater improvement in the HbA_{1c} of the intervention group than that of the usual care group (mean difference 1.2%; 95% CI 0.5-1.9; P=.001) at 12 months follow-up [47]. In the study with Noom, Toro-Ramos et al [46] found no difference in the HbA_{1c} (mean difference 0.006%; SE 0.07; P=.93) between the intervention and control groups at 12 months follow-up [46]. Johansen et al [46] found that the Liva intervention did not reach the prespecified criterion for equivalence (mean difference −0.26%; 95% CI −0.52 to −0.01; P=.15) [46]. In the randomized crossover trial of Livongo, Amante et al [49] reported similar rates of HbA_{1c} change in both groups (intervention/usual care and usual care/intervention), and a significant treatment effect (mean change for intervention/usual care −1.1%, SD 1.5; mean change for usual care/intervention −0.8%, SD 1.5; P<.001) during the first 6 months. However, in the mixed-effects model, there was no significant improvement in HbA_{1c} between the intervention and usual care conditions (mean change 0.4%; P=.06). Compared to baseline, the interventions of Noom, Liva, and BlueStar showed HbA_{1c} reductions of 0.23% [45], 0.31% [46], and 1.9% [47] at 12 months, respectively. Using Livongo yielded HbA_{1c} reductions of 0.9% and 1.2% for the intervention/usual care and usual care/intervention group, respectively [48]. A summary of all the reported effectiveness measures among the identified scientific publications can be found in Multimedia Appendix 3.

Characteristics of DBCIs

The full list of the included DBCIs is outlined in Table 3. Overall, 11 DBCIs were found to address diabetes prevention and management (Noom, Omada, Livongo, Vida, Lark, BlueStar, Liva, Oviva, GlycoLeap, Sweetch, and BEATdiabetes), whereas 4 DBCIs addressed only diabetes management (Virta, Dario, One Drop, and Twin), and 1 solely focused on diabetes prevention (Fruit Street). The program costs varied, ranging from US $19.99 to $249 per month, whereas some were available on an annual basis or covered by health care providers, health plans, or employers. Furthermore, 11 DBCIs (Noom, Virta, Omada, Vida, One Drop, BlueStar, Liva, Oviva, GlycoLeap, BEATdiabetes, and Fruit Street) involved a human health professional as part of the intervention delivery, and 2 DBCIs (Livongo and Dario) offered it as an optional feature. Among the 3 remaining DBCIs, 2 did not employ a health professional (Lark and Sweetch), and this could not be determined in 1 DBCI (Twin). Of the 16 included DBCIs, 10 apps were accessible to the authors. Only 1 of the 10 accessible apps employed a CA (Lark).

We found that all the 10 accessible apps (10/16, 63%) tracked health or behavioral outcomes using self-reports as well as sensor and device analytics. Diet and body weight were the most frequently tracked health and behavioral outcomes (n=10), followed by physical activity or exercise (n=9), blood glucose (n=7), blood pressure and HbA_{1c} (n=5), mood (n=3), sleep (n=3), medication (n=2), waist circumference (n=1), well-being (n=1), calories (n=1), heart rate (n=1), and stress (n=1).

Table 3. Intervention delivery characteristics of the companies' digital behavior change interventions.

DBCI[a] name	Health continuum category	Cost	HHP[b] involved	CA[c] used	Tracked health and behavioral outcomes	Self-reports; sensor and device analytics
Noom	Prevention and management	US $59/month or $199/year	Yes	No	Physical activity, body weight, sleep, diet, and blood pressure	Open questions, ratings, multiple choice, physical activity recordings, and accelerometer gyroscope
Virta	Management	US $249/month plus a one-time $250 initiation fee	Yes	—[d]	—	—
Omada	Prevention and management	US $140/month for the first 4 months and $20/month for the following months	Yes	No	Blood glucose, physical activity, body weight, diet, and blood pressure	Open questions, ratings, multiple choice, body sensors, physical activity recordings, and Bluetooth
Livongo	Prevention and management	Purchase free; costs covered by employer, health plan, or health care provider	Yes, but optional	No	HbA_{1c}[e], blood glucose, physical activity, body weight, diet, and blood pressure	Open questions, ratings, body sensors, camera, Bluetooth, and accelerometer gyroscope
Vida	Prevention and management	Free download, free 1 week trial, and subscription US $58.25-$79/month	Yes	No	HbA_{1c}, physical activity, body weight, stress, and diet	Open questions, ratings, multiple choice, and Bluetooth
Dario	Management	Basic US $25-$30/month, pro US $33-$40/month, and premium US $70-$85/ month	Yes, but optional	—	—	—
One Drop	Management	Digital membership US $19.99/month, supplies $20.99/month, and combined package $30.99/month	Yes	No	HbA_{1c}, blood glucose, physical activity, body weight, medication, diet, and blood pressure	Open questions, ratings, multiple choice, location, camera, and telephone
Lark	Prevention and management	Lark Weight Loss Pro US $19.99, Lark Wellness Pro $14.99, and Lark Diabetes Prevention Program Pro $119.99	No	Yes	Physical activity, body weight, sleep, mood, well-being, and diet	Open questions, ratings, multiple choice, Bluetooth, accelerometer gyroscope, GPS[f], and app usage
BlueStar	Prevention and management	Unclear	Yes	—	—	—
Liva	Prevention, Management	Unclear	Yes	—	—	—
Twin	Management	INR[g] 1450 (1 INR=US $0.01344) for a 14-day trial; price for continuous use unclear	Unclear	—	—	—
Oviva	Prevention and management	CHF[h] 484 (1 CHF=US $1.09204) carried by health care provider	Yes	No	Blood glucose, physical activity, body weight, mood, and diet	Open questions, ratings, multiple choice, and camera
GlycoLeap	Prevention and management	Free, but only available for diabetic and prediabetic patients through their doctor if they are part of the project or through particular employers	Yes	No	HbA_{1c}, blood glucose, body weight, mood, and diet	Open questions, ratings, camera, Bluetooth, and photos

DBCI[a] name	Health continuum category	Cost	HHP[b] involved	CA[c] used	Tracked health and behavioral outcomes	Self-reports; sensor and device analytics
Sweetch	Prevention and management	Unclear	No	—	—	—
BEATdiabetes	Prevention and Management	Unclear	Yes	No	HbA_{1c}, blood glucose, physical activity, body weight, medication, waist circumference, and diet	Open questions, ratings, multiple choice, and Bluetooth
Fruit Street	Prevention	US $19.99/month	Yes	No	Blood glucose, physical activity, body weight, sleep, heart rate, calories, diet, and blood pressure	Open questions, ratings, physical activity recordings, camera, and Bluetooth

[a]DBCI: digital behavior change intervention.

[b]HHP: human health professional.

[c]CA: conversational agent.

[d]—app not accessible.

[e]HbA_{1c}: glycated hemoglobin A_{1c}.

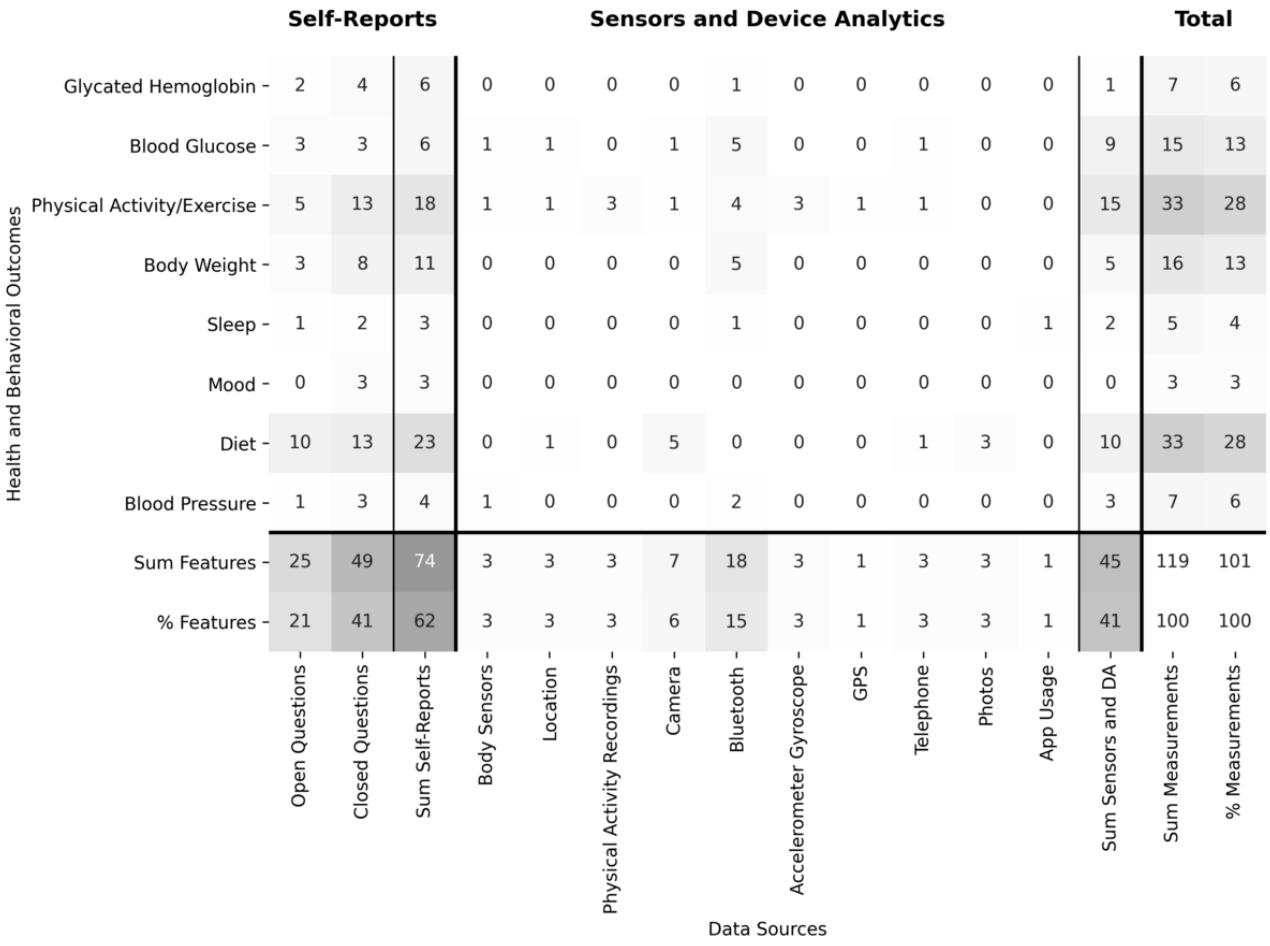
[f]GPS: Global Positioning System.

[g]INR: Indian Rupee.

[h]CHF: Swiss Franc.

The findings regarding the usage of self-reports as well as sensor and device analytics are summarized in Figure 2. In the 119 usages considered, self-reports were used 74 times (62%), whereas sensor and device analytics were used 45 times (38%) as the data source of the 10 accessible apps. Self-reports were most frequently measured by closed questions including ratings, Likert scales, and multiple-choice questions (49 times, 41%) followed by open questions (25 times, 21%). The sensor and device analytics that were most frequently used were Bluetooth and cameras, which were used 18 (15%) and 7 times (6%), respectively. The darker color indicates a higher number of occurrences.

Figure 2. Gray scale illustrating the number of times health or behavioral outcomes were measured by self-reports or sensor and device analytics summarized considering all the 10 reviewed apps. DA: device analytics; GPS: Global Positioning System.



Discussion

Principal Results

Of the 16 companies and DBCIs included in this review, only 4 were assessed for their effectiveness in changing HbA_{1c} via high-quality RCTs. Results from the 4 RCTs analyzed indicate these DBCIs have a varying effect on HbA_{1c}. For example, the BlueStar intervention showed a significant improvement of 1.2% in HbA_{1c} compared to the usual care group at 12 months, whereas the Noom, Livongo, and Liva interventions did not show any significant improvements. Furthermore, there was a wide range in the number of effectiveness studies across DBCIs, with 1 study having no published scientific evidence to 1 having 11 associated publications. We found a trend toward more published studies involving higher-funded companies, with the 3 top-funded companies (Noom, Virta, and Omada) accounting for more than half (26/50, 52%) of all publications. We also found that 5 of the highest-funded DBCIs achieved full recognition status from the DPRP (Noom, Omada, Livongo, Vida, and Lark), whereas only 1 among the lower-funded companies with funding ranks 9 to 16 (Fruit Street) received full DPRP recognition. Further, 2 DBCIs in our sample (Dario and BEATdiabetes) were neither recognized by the DPRP nor had any published effectiveness studies available. More adequately powered and high-quality RCTs are needed to confirm the effectiveness of top-funded DBCIs for type 2 diabetes prevention and management.

Recognition by national authorities to deliver evidence-based programs can be an important reference point for potential consumers and physicians when deciding to use or prescribe a particular intervention program and can serve to incentivize the adoption of impact-focused interventions [36]. Recognition can benefit the companies offering the interventions by providing sustainability and reimbursement for the intervention through many private and public payers that require recognition, such as Medicare [49]. Recognition can also be an effective marketing tool and encourage referrals. However, we only found 1 certification program for evidence-based diabetes prevention or management programs, which was the DPRP offered by the US CDC [43]. This lack of quality assurance programs could hamper consumers' and health care providers' decision-making processes when identifying the most effective programs. Therefore, additional quality assurance programs that can certify diabetes prevention and management interventions based on evidence-based criteria are necessary, especially for diabetes management interventions and in countries other than the United States.

Reduction in HbA_{1c} is one of the key clinical outcomes for assessing the effectiveness of interventions for type 2 diabetes prevention and management and is also one of the effectiveness criteria to achieve recognition by the DPRP [44]. In the 4 RCTs

evaluated in our analysis, the Noom and Liva interventions showed modest HbA_{1c} reductions of 0.2% to 0.3% [45,46], whereas the BlueStar and Livongo interventions showed higher HbA_{1c} reductions of over 1% [47,48]. According to the criteria of the DPRP, an HbA_{1c} reduction of 0.2 percentage points is considered sufficient for a lifestyle change program to receive recognition [44], although a change of 0.4% to 0.5% is considered a clinically meaningful improvement [50]. The 4 RCTs reviewed [46-49] were also powered to detect changes in HbA_{1c} between 0.4% and 1%. Therefore, this raises the question of whether the effectiveness criterion of the DPRP standards around the change in HbA_{1c} is sufficient. Furthermore, even though recognition from the DPRP guarantees that a certain level of diabetes risk reduction was achieved because of a specific DBCI, the recognition does not give any further information on the magnitude of the reduction, as data that companies submit to achieve DPRP recognition are not made publicly available. This lack of information limits transparency for researchers, investors, users, and payers to identify the most effective programs. Moreover, this lack of data transparency could become even more troublesome if companies that are already recognized to deliver evidence-based programs are then unwilling to invest additional resources into research and development. Therefore, we highlight the need for more transparency regarding data related to the effectiveness of DBCIs. We believe that our findings also indicate the importance of encouraging the digital health industry to build more evidence-based DBCIs. Clarifying the regulatory landscape around DBCIs and developing incentives that lead to a stronger customer market have been identified as 2 possible areas that policy makers may address to foster such an encouragement [36]. In addition, we recognize the poor standard of reporting by the DBCI companies regarding the app features, employed behavior change techniques, and information on what and how sensing data are being utilized. This lack of transparent reporting is likely because companies that develop these proprietary apps tend to be reluctant to disclose app details that could potentially be useful for competitors. From a research perspective, this lack of transparency makes it difficult to compare intervention features objectively. It also reveals the need for more transparent reporting on the characteristics of DBCIs by the companies.

In our reviewed DBCIs, the most commonly tracked health and behavioral outcomes were diet and body weight, which were tracked in all the 10 accessible apps, followed by physical activity or exercise, which was tracked in 9 apps. Other frequently tracked outcomes were blood glucose (7 apps), blood pressure, and HbA_{1c} (5 apps each). Our findings are in line with previous studies that reviewed apps for self-management and lifestyle modification in type 2 diabetes patients [51-53] and are also similar to the opinion of clinical experts regarding important intervention components [53,54]. However, we found that less than 40% of health and behavioral outcomes were measured using sensors and device analytics and that most outcomes were measured by self-reports. Although such self-reports can be used in the form of ecological momentary assessments [55] that are closely related to the concept of JITAIs [56], self-reports can be burdensome for participants to complete and may lead to difficulties in keeping users engaged [25,57]. Therefore, we believe that self-reports are not sufficient to leverage the full potential of JITAIs. The low usage of measurements from sensor and device analytics indicates that it is unlikely that the investigated interventions use JITAI mechanisms to tailor the intervention content to the user. In addition, there is no clear evidence on how these intervention components are related to intervention effectiveness; therefore, future studies must identify which DBCI features most successfully impact intervention effectiveness.

Our review also aimed to assess the extent to which human health professionals and automated CAs are used within the DBCIs. We found that 13 of the 16 DBCIs involved a human health professional, of which 2 DBCIs offered it as an optional feature. We found that among the 10 apps that were available to us, only 1 app used a CA. The high usage of human health coaches alongside the low usage of CAs, and the unlikely use of JITAI mechanisms to tailor intervention content, indicates the low use of automation among the investigated DBCIs. This limits the overall scalability of existing DBCIs and the potential of the interventions to reach a greater proportion of the eligible population [58] because the involvement of human health coaches is generally time- and resource-intensive.

We identified 4 potential reasons that might account for this low use of automation among the investigated DBCIs. First, automated approaches, such as CAs, are still part of an emerging area within type 2 diabetes management and DBCIs. It is possible that users might have concerns when relying on CAs for actionable medical information around diabetes [59]. Second, app features that use sensor technologies might not be adequately developed to replace input from human health professionals or self-report methods, thus leading to significant user burden. For example, the current state-of-the-art food volume estimation approaches to assess dietary intake are not yet usable in commercial apps due to several gaps and technological issues [60]. Therefore, many apps rely on user inputs, for example, by selecting serving sizes of identified foods, based on which nutritional values can be estimated [60]. Third, there appears to be insufficient evidence to support the widespread use of fully automated approaches without remote access to a human health professional [13]. Thus, additional RCTs or cohort studies that directly compare DBCIs involving digital human coaches with fully automated approaches are needed to better understand the potential and effectiveness of automated DBCIs. Fourth, in the current standards and operating procedures of the DPRP, live interactions with lifestyle coaches should be offered at least on a weekly basis during the first 6 months [44]. Although email and text message interactions may contribute toward this requirement, it is likely to be challenging for companies aiming to offer fully automated DBCIs to meet this requirement. Recognition by the DPRP is valuable to many companies [49]; nevertheless, satisfying the requirement of offering live coaching interactions prevents the recognition of fully automated approaches and limits the scalability of DBCIs for type 2 diabetes prevention. Further research is warranted to establish if human coaches are indeed necessary to deliver an effective lifestyle change program.

Strengths and Limitations

This study has several strengths. First, we conducted a comprehensive company search involving 2 widely used venture capital databases [35], and we had 3 independent digital health experts confirm that the final list of included companies covered the market. Second, we conducted comprehensive data extraction using multiple sources, including databases, intervention websites, peer-reviewed publications, and mobile apps. Third, we summarized only the highest quality scientific evidence on the effectiveness of the included DBCIs.

However, our review has some limitations. First, even though we identified the top-funded companies in the field, this does not guarantee that their interventions reach a significant proportion of the target population. Many of the reviewed companies are still in the start-up phase where they typically acquire considerable funding; however, their DBCIs may have limited accessibility, for example, only through referral by partnering clinicians. Second, we were only able to access 10 out of the 16 DBCI apps, as some apps were only accessible with a subscription service, in a specific geographic region, with a doctor's prescription, an access code, or through an employer subscription. Although we systematically contacted the companies and requested app access, we only received additional access to 5 paid or proprietary apps through the companies. Third, we were unable to access, and therefore assess, any additional devices that may have accompanied the DBCI apps. Some of these devices record additional health parameters via sensors, such as (smart) blood glucose meters, (smart or wireless) scales, activity trackers, or smartwatches. Therefore, we could not assess all the functionalities of these devices, which limited the comprehensiveness of our review. Fourth, we were not able to assess certain app features that were behind a paywall. This was often the case for support that was delivered through health professionals. Fifth, the DPRP is only relevant for interventions targeting diabetes prevention and therefore does not cover DBCIs that solely target diabetes management. In addition, not all reviewed DBCIs were available in the United States; consequently, they are not eligible to achieve recognition by the DPRP. Sixth, given that most of the investigated DBCIs and all DBCIs with a corresponding fully powered RCT address diabetes management and diabetes prevention, it was not feasible to separately report the results in these 2 categories.

Comparison With Prior Work

This is the first systematic assessment of the top-funded companies that offer DBCIs for type 2 diabetes prevention or management. Previous reviews have focused on apps and digital interventions for diabetes management, but they were mostly limited to interventions reported in scientific research without a particular impact on the market [51,52,61-64]. These reviews generally found DBCIs to be effective in improving diabetes-related outcomes, particularly HbA_{1c} [51,52,61-64], which is in line with our findings; nevertheless, they also concur that the current evidence is limited and there is a need for adequately powered, rigorous trials with long-term follow-ups to determine the clinical and economic impact of such interventions [52,65]. In terms of JITAIs, a recent systematic review investigating popular mental health apps for individuals with depression concluded that JITAI mechanisms have not yet been translated into mainstream depression apps [66], which also aligns with our findings.

Conclusions

Our findings suggest that the level of funding received by companies offering DBCIs for type 2 diabetes prevention and management does not coincide with the level of evidence on the intervention effectiveness. There is significant variation in the level of evidence underpinning the different DBCIs and an overall need for more rigorous effectiveness trials as well as additional certification programs for evidence-based diabetes prevention and management interventions in countries other than the United States. In addition, we emphasize the need for more data transparency from quality assurance authorities to inform stakeholders and consumers on how effective each DBCI is in improving diabetes-related outcomes. We further found low usage of CAs, an unlikely use of JITAI mechanisms, and a high level of support from human health professionals among the apps investigated, which indicates low usage of automated approaches. Because automation and technology are critical factors to increase the interventions' scalability, further research is warranted to establish the effectiveness of fully automated DBCIs in comparison to those offering support from human health professionals. Finally, we recommend that national authorities such as the DPRP help reduce barriers for the recognition of fully automated approaches and encourage policy makers to foster an environment that encourages the digital health industry to build more evidence-based solutions.

Acknowledgments

We would like to thank Dr Alicia Salamanca-Sanabria for reviewing the manuscript and providing feedback. The research was conducted as part of the Future Health Technologies program, which was established collaboratively between ETH Zurich and the National Research Foundation, Singapore. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise program and by CSS Insurance (Switzerland).

Authors' Contributions

RK, SH, KL, and TK were responsible for the study design. GWT and TK developed the underlying framework defining which areas to investigate and how to review these areas. RK adapted this framework for the study. SH and KL were responsible for the search strategy. RK and SH were responsible for screening and data extraction. RK and GWT were responsible for aggregation,

cleaning, and condensation of the data. RK was responsible for the first draft. All authors were responsible for critical feedback and final revisions to the manuscript. JM and TK share last authorship.

Conflicts of Interest

SH, GWT, KL, FvW, and TK are affiliated with the Center for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St Gallen, which is funded in part by CSS, a Swiss health insurer. TK is also the cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, Pathmate Technologies was not involved in any way in the design, interpretation, and analysis during the study, or in writing the paper.

Multimedia Appendix 1
Codebook for app review.
[DOCX File , 23 KB - jmir_v24i1e33348_app1.docx]

Multimedia Appendix 2
Company and intervention characteristics.
[PDF File (Adobe PDF File), 177 KB - jmir_v24i1e33348_app2.pdf]

Multimedia Appendix 3
Study characteristics.
[PDF File (Adobe PDF File), 122 KB - jmir_v24i1e33348_app3.pdf]

References

1. International DF. IDF Diabetes Atlas 9th Edition. Brussels, Belgium: International Diabetes Federation; 2019.
2. Global Report on Diabetes. Geneva, Switzerland: World Health Organisation; 2016.
3. Khan MAB, Hashim MJ, King JK, Govender RD, Mustafa H, Al Kaabi J. Epidemiology of type 2 diabetes—global burden of disease and forecasted trends. J Epidemiol Glob Health 2020 Mar;10(1):107-111 [FREE Full text] [doi: 10.2991/jegh.k.191028.001] [Medline: 32175717]
4. Nathan DM, Davidson MB, DeFronzo RA, Heine RJ, Henry RR, Pratley R, American Diabetes Association. Impaired fasting glucose and impaired glucose tolerance: implications for care. Diabetes Care 2007 Mar;30(3):753-759. [doi: 10.2337/dc07-9920] [Medline: 17327355]
5. American Diabetes Association. 5. Facilitating behavior change and well-being to improve health outcomes: standards of medical care in diabetes—2020. Diabetes Care 2020 Jan;43(Suppl 1):S48-S65 [FREE Full text] [doi: 10.2337/dc20-S005] [Medline: 31862748]
6. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care 2002 Dec;25(12):2165-2171 [FREE Full text] [doi: 10.2337/diacare.25.12.2165] [Medline: 12453955]
7. Penn L, Rodrigues A, Haste A, Marques MM, Budig K, Sainsbury K, et al. NHS Diabetes Prevention Programme in England: formative evaluation of the programme in early phase implementation. BMJ Open 2018 Feb;8(2):e019467 [FREE Full text] [doi: 10.1136/bmjopen-2017-019467] [Medline: 29467134]
8. Horigan G, Davies M, Findlay-White F, Chaney D, Coates V. Reasons why patients referred to diabetes education programmes choose not to attend: a systematic review. Diabet Med 2017 Jan;34(1):14-26. [doi: 10.1111/dme.13120] [Medline: 26996982]
9. Azar KMJ, Koliwad S, Poon T, Xiao L, Lv N, Griggs R, et al. The Electronic CardioMetabolic Program (eCMP) for patients with cardiometabolic risk: a randomized controlled trial. J Med Internet Res 2016 May;18(5):e134 [FREE Full text] [doi: 10.2196/jmir.5143] [Medline: 27234480]
10. Smith KJ, Kuo S, Zgibor JC, McTigue KM, Hess R, Bhargava T, et al. Cost effectiveness of an internet-delivered lifestyle intervention in primary care patients with high cardiovascular risk. Prev Med 2016 Jun;87:103-109. [doi: 10.1016/j.ypmed.2016.02.036] [Medline: 26921656]
11. Gamble A, Pham Q, Goyal S, Cafazzo JA. The challenges of COVID-19 for people living with diabetes: considerations for digital health. JMIR Diabetes 2020 May;5(2):e19581 [FREE Full text] [doi: 10.2196/19581] [Medline: 32392473]
12. Hekler EB, Michie S, Pavel M, Rivera DE, Collins LM, Jimison HB, et al. Advancing models and theories for digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):825-832. [doi: 10.1016/j.amepre.2016.06.013] [Medline: 27745682]
13. Tice J, Chapman R, Shore K, Seidner M, Ollendorf D, Weissberg J, et al. Diabetes prevention programs: effectiveness and value; final evidence report and meeting summary. Institute for Clinical and Economic Review (ICER). URL: https://icer-review.org/material/final-report-dpp/ [accessed 2021-02-03]
14. Yardley L, Choudhury T, Patrick K, Michie S. Current issues and future directions for research into digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):814-815. [doi: 10.1016/j.amepre.2016.07.019] [Medline: 27745680]

15. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
16. Bérubé C, Schachner T, Keller R, Fleisch E, V Wangenheim F, Barata F, et al. Voice-based conversational agents for the prevention and management of chronic and mental health conditions: systematic literature review. J Med Internet Res 2021 Mar;23(3):e25933 [FREE Full text] [doi: 10.2196/25933] [Medline: 33658174]
17. Schachner T, Keller R, V Wangenheim F. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. J Med Internet Res 2020 Sep;22(9):e20701 [FREE Full text] [doi: 10.2196/20701] [Medline: 32924957]
18. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
19. Bickmore TW, Mitchell SE, Jack BW, Paasche-Orlow MK, Pfeifer LM, Odonnell J. Response to a relational agent by hospital patients with depressive symptoms. Interact Comput 2010 Jul;22(4):289-298 [FREE Full text] [doi: 10.1016/j.intcom.2009.12.001] [Medline: 20628581]
20. Ma T, Chattopadhyay D, Sharifi H. Virtual humans in health-related interventions: a meta-analysis. In: CHI EA '19: Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI Conference on Human Factors in Computing Systems; May 4-9, 2019; Glasgow, Scotland p. 1-6. [doi: 10.1145/3290607.3312853]
21. Bickmore TW, Picard RW. Establishing and maintaining long-term human-computer relationships. ACM Trans Comput Hum Interact 2005 Jun;12(2):293-327. [doi: 10.1145/1067860.1067867]
22. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
23. Kowatsch T, Schachner T, Harperink S, Barata F, Dittler U, Xiao G, et al. Conversational agents as mediating social actors in chronic disease management involving health care professionals, patients, and family members: multisite single-arm feasibility study. J Med Internet Res 2021 Feb;23(2):e25060 [FREE Full text] [doi: 10.2196/25060] [Medline: 33484114]
24. Mitchell EG, Maimone R, Cassells A, Tobin JN, Davidson P, Smaldone AM, et al. Automated vs. human health coaching. Proc ACM Hum-Comput Interact 2021 Apr;5(CSCW1):1-37. [doi: 10.1145/3449173]
25. Nahum-Shani I, Smith SN, Spring BJ, Collins LM, Witkiewitz K, Tewari A, et al. Just-in-time adaptive interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2016 Sep;52(6):446-462. [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
26. Nahum-Shani I, Hekler EB, Spruijt-Metz D. Building health behavior models to guide the development of just-in-time adaptive interventions: a pragmatic framework. Health Psychol 2015 Dec;34(Suppl):1209-1219 [FREE Full text] [doi: 10.1037/hea0000306] [Medline: 26651462]
27. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr;16(1):1-21 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
28. Spruijt-Metz D, Wen CKF, O'Reilly G, Li M, Lee S, Emken BA, et al. Innovations in the use of interactive technology to support weight management. Curr Obes Rep 2015 Dec;4(4):510-519 [FREE Full text] [doi: 10.1007/s13679-015-0183-6] [Medline: 26364308]
29. Thomas JG, Bond DS. Behavioral response to a just-in-time adaptive intervention (JITAI) to reduce sedentary behavior in obese adults: implications for JITAI optimization. Health Psychol 2015 Dec;34:1261-1276 [FREE Full text] [doi: 10.1037/hea0000304] [Medline: 26651467]
30. Fleming GA, Petrie JR, Bergenstal RM, Holl RW, Peters AL, Heinemann L. Diabetes digital app technology: benefits, challenges, and recommendations. A consensus report by the European Association for the Study of Diabetes (EASD) and the American Diabetes Association (ADA) Diabetes Technology Working Group. Diabetes Care 2020 Jan;43(1):250-260. [doi: 10.2337/dci19-0062] [Medline: 31806649]
31. Digital diabetes care market 2020: going beyond diabetes management. Research2Guidance. Berlin, Germany: Research2Guidance; 2020. URL: https://research2guidance.com/product/the-global-digital-diabetes-care-market-2020-going-beyond-diabetes-management/ [accessed 2022-01-04]
32. Klonoff DC, Evans B, Zweig M, Day S, Kerr D. Is digital health for diabetes in an investment bubble? J Diabetes Sci Technol 2020 Jan;14(1):165-169. [doi: 10.1177/1932296819867742] [Medline: 31470739]
33. Crunchbase. URL: https://www.crunchbase.com/ [accessed 2020-07-23]
34. PitchBook. URL: https://pitchbook.com/ [accessed 2020-07-23]
35. Retterath A, Braun R. Benchmarking venture capital databases. SSRN Journal 2020 Nov:1-47 [FREE Full text] [doi: 10.2139/ssrn.3706108]
36. Safavi K, Mathews SC, Bates DW, Dorsey ER, Cohen AB. Top-funded digital health companies and their impact on high-burden, high-cost conditions. Health Aff (Millwood) 2019 Jan;38(1):115-123. [doi: 10.1377/hlthaff.2018.05081] [Medline: 30615535]

37. Smartphone market share. IDC. URL: https://www.idc.com/promo/smartphone-market-share/os [accessed 2021-12-22]
38. Covidence Systematic Review Software. Melbourne, Australia: Veritas Health Innovation URL: https://www.covidence.org/ [accessed 2020-07-20]
39. Procedure Manual Section 4. Evidence Review Development. United States: U.S. Preventive Services Task Force (USPSTF) URL: https://uspreventiveservicestaskforce.org/uspstf/about-uspstf/methods-and-processes/procedure-manual/procedure-manual-section-4-evidence-review-development [accessed 2020-11-29]
40. Cornet VP, Holden RJ. Systematic review of smartphone-based passive sensing for health and wellbeing. J Biomed Inform 2018 Jan;77:120-132. [doi: 10.1016/j.jbi.2017.12.008] [Medline: 29248628]
41. Sensors overview. Developers.: Developers Android URL: https://developer.android.com/guide/topics/sensors/sensors_overview [accessed 2020-07-07]
42. ioS security guide. Apple Platform Security. URL: https://support.apple.com/en-gb/guide/security/welcome/web [accessed 2020-07-07]
43. National Diabetes Prevention Program. United States: Centers for Disease Control and Prevention URL: https://www.cdc.gov/diabetes/prevention/index.html [accessed 2021-08-15]
44. Diabetes Prevention Recognition Program: Standards and Operating Procedures. United States: Centers for Disease Control and Prevention URL: https://nationaldppcsc.cdc.gov/s/article/DPRP-Standards-and-Operating-Procedures [accessed 2021-05-25]
45. Toro-Ramos T, Michaelides A, Anton M, Karim Z, Kang-Oh L, Argyrou C, et al. Mobile delivery of the diabetes prevention program in people with prediabetes: randomized controlled trial. JMIR Mhealth Uhealth 2020 Jul;8(7):e17842 [FREE Full text] [doi: 10.2196/17842] [Medline: 32459631]
46. Johansen MY, MacDonald CS, Hansen KB, Karstoft K, Christensen R, Pedersen M, et al. Effect of an intensive lifestyle intervention on glycemic control in patients with type 2 diabetes: a randomized clinical trial. JAMA 2017 Dec;318(7):637-646. [doi: 10.1001/jama.2017.10169] [Medline: 28810024]
47. Quinn CC, Shardell MD, Terrin ML, Barr EA, Ballew SH, Gruber-Baldini AL. Cluster-randomized trial of a mobile phone personalized behavioral intervention for blood glucose control. Diabetes Care 2011 Sep;34(9):1934-1942 [FREE Full text] [doi: 10.2337/dc11-0366] [Medline: 21788632]
48. Amante DJ, Harlan DM, Lemon SC, McManus DD, Olaitan OO, Pagoto SL, et al. Evaluation of a diabetes remote monitoring program facilitated by connected glucose meters for patients with poorly controlled type 2 diabetes: randomized crossover trial. JMIR Diabetes 2021 Mar;6(1):e25574 [FREE Full text] [doi: 10.2196/25574] [Medline: 33704077]
49. FAQ:CDC's Diabetes Prevention Recognition Program. United States: Centers for Disease Control and Prevention URL: https://nationaldppcsc.cdc.gov/s/article/FAQ-CDCs-Diabetes-Prevention-Recognition-Program-1526419472100 [accessed 2021-08-15]
50. Little RR, Rohlfing CL. The long and winding road to optimal HbA1c measurement. Clin Chim Acta 2013 Mar;418:63-71 [FREE Full text] [doi: 10.1016/j.cca.2012.12.026] [Medline: 23318564]
51. Wu X, Guo X, Zhang Z. The efficacy of mobile phone apps for lifestyle modification in diabetes: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Jan;7(1):e12297 [FREE Full text] [doi: 10.2196/12297] [Medline: 30664494]
52. Veazie S, Winchell K, Gilbert J, Paynter R, Ivlev I, Eden KB, et al. Rapid evidence review of mobile applications for self-management of diabetes. J Gen Intern Med 2018 Jul;33(7):1167-1176. [doi: 10.1007/s11606-018-4410-1] [Medline: 29740786]
53. Mehraeen E, Noori T, Nazeri Z, Heydari M, Mehranfar A, Moghaddam HR, et al. Identifying features of a mobile-based application for self-care of people living with T2DM. Diabetes Res Clin Pract 2021 Jan;171:108544. [doi: 10.1016/j.diabres.2020.108544] [Medline: 33227362]
54. Salari R, Niakan Kalhori SR, Fatehi F, Ghazisaeedi M, Nazari M. Determining minimum set of features for diabetes mobile apps. J Diabetes Metab Disord 2019 Dec;18(2):333-340 [FREE Full text] [doi: 10.1007/s40200-019-00417-y] [Medline: 31890658]
55. Shiffman S, Stone AA, Hufford MR. Ecological momentary assessment. Annu Rev Clin Psychol 2008 Apr;4:1-32. [doi: 10.1146/annurev.clinpsy.3.022806.091415] [Medline: 18509902]
56. Wang L, Miller LC. Just-in-the-moment adaptive interventions (JITAI): a meta-analytical review. Health Commun 2019 Sep:1-14. [doi: 10.1080/10410236.2019.1652388] [Medline: 31488002]
57. Sharmin M, Raij A, Epstien D, Nahum-Shani I, Beck JG, Vhaduri S, et al. Visualization of time-series sensor data to inform the design of just-in-time adaptive stress interventions. Proc ACM Int Conf Ubiquitous Comput 2015 Sep;2015:505-516 [FREE Full text] [doi: 10.1145/2750858.2807537] [Medline: 26539566]
58. Milat AJ, King L, Bauman AE, Redman S. The concept of scalability: increasing the scale and potential adoption of health promotion interventions into policy and practice. Health Promot Int 2013 Sep;28(3):285-298 [FREE Full text] [doi: 10.1093/heapro/dar097] [Medline: 22241853]
59. Bickmore TW, Trinh H, Olafsson S, O'Leary TK, Asadi R, Rickles NM, et al. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. J Med Internet Res 2018 Sep;20(9):e11510 [FREE Full text] [doi: 10.2196/11510] [Medline: 30181110]

60. Tay W, Kaur B, Quek R, Lim J, Henry CJ. Current developments in digital quantitative volume estimation for the optimisation of dietary assessment. Nutrients 2020 Apr;12(4):1167 [FREE Full text] [doi: 10.3390/nu12041167] [Medline: 32331262]
61. Bonoto BC, de Araújo VE, Godói IP, de Lemos LLP, Godman B, Bennie M, et al. Efficacy of mobile apps to support the care of patients with diabetes mellitus: a systematic review and meta-analysis of randomized controlled trials. JMIR Mhealth Uhealth 2017 Mar;5(3):e4 [FREE Full text] [doi: 10.2196/mhealth.6309] [Medline: 28249834]
62. Hou C, Carter B, Hewitt J, Francisa T, Mayor S. Do mobile phone applications improve glycemic control (HbA1c) in the self-management of diabetes? A systematic review, meta-analysis, and GRADE of 14 randomized trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dc16-0346] [Medline: 27926892]
63. Kebede MM, Peters M, Heise TL, Pischke CR. Comparison of three meta-analytic methods using data from digital interventions on type 2 diabetes. Diabetes Metab Syndr Obes 2018 Dec;12:59-73 [FREE Full text] [doi: 10.2147/DMSO.S180106] [Medline: 30588055]
64. Kebede MM, Zeeb H, Peters M, Heise TL, Pischke CR. Effectiveness of digital interventions for improving glycemic control in persons with poorly controlled type 2 diabetes: a systematic review, meta-analysis, and meta-regression analysis. Diabetes Technol Ther 2018 Nov;20(11):767-782. [doi: 10.1089/dia.2018.0216] [Medline: 30257102]
65. Drincic A, Prahalad P, Greenwood D, Klonoff DC. Evidence-based mobile medical applications in diabetes. Endocrinol Metab Clin North Am 2016 Dec;45(4):943-965 [FREE Full text] [doi: 10.1016/j.ecl.2016.06.001] [Medline: 27823614]
66. Teepe GW, Da Fonseca A, Kleim B, Jacobson NC, Salamanca Sanabria A, Tudor Car L, et al. Just-in-time adaptive mechanisms of popular mobile apps for individuals with depression: systematic app search and literature review. J Med Internet Res 2021 Sep;23(9):e29412 [FREE Full text] [doi: 10.2196/29412] [Medline: 34309569]

Abbreviations

CAs: conversational agents
CDC: Center for Disease Control and Prevention
DBCIs: digital behavior change interventions
DPRP: Diabetes Prevention Recognition Program
HbA_{1c}: glycated hemoglobin A1c
JTAI: just-in-time adaptive intervention
RCTs: randomized controlled trials

Edited by G Eysenbach; submitted 06.09.21; peer-reviewed by K Pal; comments to author 28.09.21; revised version received 22.10.21; accepted 15.11.21; published 07.01.22.

Please cite as:
Keller R, Hartmann S, Teepe GW, Lohse KM, Alattas A, Tudor Car L, Müller-Riemenschneider F, von Wangenheim F, Mair JL, Kowatsch T
Digital Behavior Change Interventions for the Prevention and Management of Type 2 Diabetes: Systematic Market Analysis
J Med Internet Res 2022;24(1):e33348
URL: https://www.jmir.org/2022/1/e33348
doi:10.2196/33348
PMID:34994693

Review

Technology Acceptance of Home-Based Cardiac Telerehabilitation Programs in Patients With Coronary Heart Disease: Systematic Scoping Review

Hadassah Joann Ramachandran[1], BSc (Hons); Ying Jiang[1], PhD; Jun Yi Claire Teo[1], BSc (Hons); Tee Joo Yeo[2], MRCP, MMED, MBBS; Wenru Wang[1], PhD

[1]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore
[2]Cardiac Rehabilitation, Department of Cardiology, National University Heart Centre, Singapore, Singapore

Corresponding Author:
Wenru Wang, PhD
Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine
National University of Singapore
10 Medical Drive, Blk MD11, Level 2
Singapore, 117597
Singapore
Phone: 65 66011761
Email: nurww@nus.edu.sg

Abstract

Background: An understanding of the technology acceptance of home-based cardiac telerehabilitation programs is paramount if they are to be designed and delivered to target the needs and preferences of patients with coronary heart disease; however, the current state of technology acceptance of home-based cardiac telerehabilitation has not been systematically evaluated in the literature.

Objective: We aimed to provide a comprehensive summary of home-based cardiac telerehabilitation technology acceptance in terms of (1) the timing and approaches used and (2) patients' perspectives on its usability, utility, acceptability, acceptance, and external variables.

Methods: We searched PubMed, CENTRAL, Embase, CINAHL, PsycINFO, and Scopus (inception to July 2021) for English-language papers that reported empirical evidence on the technology acceptance of early-phase home-based cardiac telerehabilitation in patients with coronary heart disease. Content analysis was undertaken.

Results: The search identified 1798 studies, of which 18 studies, with 14 unique home-based cardiac telerehabilitation programs, met eligibility criteria. Technology acceptance (of the home-based cardiac telerehabilitation programs) was mostly evaluated at intra- and posttrial stages using questionnaires (n=10) and usage data (n=11). The least used approach was evaluation through qualitative interviews (n=3). Usability, utility, acceptability, and acceptance were generally favored. External variables that influenced home-based cardiac telerehabilitation usage included component quality, system quality, facilitating conditions, and intrinsic factors.

Conclusions: Home-based cardiac telerehabilitation usability, utility, acceptability, and acceptance were high; yet, a number of external variables influenced acceptance. Findings and recommendations from this review can provide guidance for developing and evaluating patient-centered home-based cardiac telerehabilitation programs to stakeholders and clinicians.

(J Med Internet Res 2022;24(1):e34657) doi:10.2196/34657

KEYWORDS

technology acceptance; coronary heart disease; home-based; telerehabilitation; web-based; mobile application; acceptance; heart; rehabilitation; app; review; evaluation; cardiac; cardiology; perspective; usability; acceptability

Introduction

Within the spectrum of cardiovascular diseases, coronary heart disease is the most common cause of mortality and morbidity globally and presents a major health care burden [1]. Cardiac rehabilitation is a widely accepted treatment modality for secondary prevention of coronary heart disease [2], but long-standing challenges regarding accessibility to cardiac rehabilitation facilities, conflicting work and care responsibilities, low socioeconomic status, and costs of rehabilitation programs have led to disappointingly low reported uptake rates among eligible patients worldwide (10% to 30% [3]). A recent challenge is the COVID-19 pandemic [4]. In the acute phase of the pandemic, nonurgent outpatient services, such as center-based cardiac rehabilitation, were partially or completely closed as limited resources and personnel were redirected to critical areas. Even in the long-term phase of the pandemic, efforts to limit the spread of COVID-19 infection through measures such as safe distancing further limited the capacity for delivery of center-based cardiac rehabilitation group exercise and therapy sessions [5]. Thus, alternative secondary prevention strategies for coronary heart disease are a priority across health care systems during the COVID-19 pandemic and beyond [4].

Home-based cardiac telerehabilitation—defined as the use of information and communication technologies (eg, mobile- and web-based platforms, wearable sensor devices) to deliver remote exercise supervision, education, counseling on cardiovascular risk factor modification, and psychosocial support exclusively at home—is one such emerging alternative [6]. A recent systematic review and meta-analysis [7] of randomized controlled trials comparing home-based cardiac telerehabilitation to center-based cardiac rehabilitation in patients with coronary heart disease found equivalent effects on functional capacity, cardiac-related hospitalization, physiological risk factor control, quality of life, depression, and behaviors such as physical activity, smoking cessation, and medication adherence. However, adapting digital solutions for health problems is not without its challenges; attempts to scale up effective digital health research interventions into real-world health care systems have been met with difficulty, especially for complex interventions that require user interaction [8,9]. The successful incorporation of such digital health technologies into clinical practice is contingent upon end-users' (ie, patients) acceptance and sustained engagement with the intervention, and thus, these are important aspect for researchers, health care systems, and policymakers to consider [9,10].

The technology acceptance model provides a framework for modeling end-user acceptance [11] and theorizes that both perceived usefulness (ie, utility) and perceived ease of use (ie, usability) of a target system directly influence intention to use (ie, acceptability), which then influences actual system use (ie, acceptance of the system) [11]. External variables such as technology self-efficacy and training, objective system design features, and the process of system implementation are thought to indirectly influence system acceptability and acceptance by influencing system utility and usability [11]. An understanding of the usability and utility of home-based cardiac telerehabilitation programs is paramount if they are to be designed and delivered to target the needs of patients with coronary heart disease in a way that ensures programs are accepted. However, the current state of technology acceptance of home-based cardiac telerehabilitation has not been systematically evaluated in the literature. We aimed to provide a comprehensive summary of the technology acceptance of home-based cardiac telerehabilitation among patients with coronary heart disease.

Methods

Study Design

We performed a systematic scoping review to comprehensively collate, summarize, and map [12,13] existing evidence on home-based cardiac telerehabilitation research in terms of usability, utility, acceptability, and acceptance testing. We used the Arksey and O'Malley methodological framework [12]: identifying the research questions, identifying relevant studies, study selection, charting the data, collating, summarizing, and reporting the results. To ensure quality and transparency, this review was conducted and reported in accordance with Preferred Reporting Items for Systematic Reviews and Meta-analyses Scoping Review guidelines [14]. A review protocol (not registered) was prepared prior to the start of this review.

Identifying the Research Question

The following research questions were identified to answer the objective of this review: (1) What are the timing and approaches used to evaluate the technology acceptance attributes in home-based cardiac telerehabilitation? (2) What are patients' perspectives on the technology acceptance constructs (ie, usability, utility, acceptability, acceptance, and external variables) of home-based cardiac telerehabilitation?

Identifying Relevant Studies

We followed recommendations by Arksey and O'Malley [12] and undertook an iterative approach, through ongoing consultations with a university resource librarian throughout the search process, to identify relevant literature. We piloted an initial search strategy in PubMed and EMBASE to identify a sample of relevant papers. This was followed by an analysis of the keywords used in the titles and abstracts and in the indexing of these relevant papers. Preliminary results revealed that terms related to the concept *acceptance* were not commonly indexed in relevant papers. Thereafter, we used terms related to coronary heart disease, rehabilitation, and telehealth. We searched PubMed, Cochrane Central Register of Controlled Trials, Embase, Cumulative Index to Nursing and Allied Health Literature, PsycINFO, and Scopus databases (inception to July 2021). No limits on study design were placed. Additionally, we manually searched the reference lists of relevant systematic reviews and papers included in this review (Table S1 in Multimedia Appendix 1).

Study Selection

Overview

Literature evaluating the technology acceptance constructs of home-based cardiac telerehabilitation that used empirical

methods (both quantitative and qualitative) and were published in English were considered. Case reports, conference abstracts, editorials, protocols, and reviews were excluded. The PCC (Population, Concept, Context) framework [14] was used to develop and set the inclusion and exclusion criteria. Search results were imported to Endnote (version X9, Clarivate Analytics) for management. Two independent authors were involved in the study selection process. Records deemed relevant by both authors were included. Consultation with a third author was used to resolve any disagreements regarding inclusion.

Population

Papers with a study population of patients with a documented medical diagnosis of coronary heart disease, acute coronary syndrome, myocardial infarction, angina pectoris, or who had undergone revascularization (ie, coronary artery bypass grafting or percutaneous coronary intervention) were included. We excluded papers with a study population of patients with heart failure (regardless of left ventricular ejection fraction), as their therapeutic needs and subsequent evaluations of home-based cardiac telerehabilitation in terms of usability, utility, acceptability, acceptance, and external variables would differ considerably from those of patients with coronary heart disease.

Concept

For the purpose of this study, the constructs of the technology acceptance model were conceptualized as follows: (1) usability—degree to which the system is easy to use and free of effort; (2) utility—degree to which the system improves user's performance and functions as intended; (3) acceptability—behavioral intention or willingness to use the system; and (4) acceptance—actual usage of the system [11,15]. Home-based cardiac telerehabilitation was defined as any mobile health app or website used either as a stand-alone platform or supplemented with other modes of delivery, such as telephone or video calls, short message service, email, or telemonitoring, to exclusively deliver early cardiac rehabilitation or secondary prevention [6]. The decision to focus on mobile- or web-based home-based cardiac telerehabilitation was made with the purpose of scoping the technologies that allowed for greater interaction, flexibility, and independence in rehabilitation programs. Papers were included if they addressed the testing and evaluation of technology acceptance constructs from patient perspectives. Late-phase home-based cardiac telerehabilitation programs, in which the focus is placed on long-term maintenance of lifestyle change, were excluded since we were only interested in the early and active rehabilitation phase (ie, focus on health behavior change, risk factor modification and psychosocial well-being.).

Context

The context for telerehabilitation programs was limited to those in a home setting only; hence, we excluded home-based cardiac telerehabilitation delivered alongside center-based cardiac rehabilitation (ie, hybrid cardiac rehabilitation services).

Charting the Data

Authors, publication year, country of origin, study design, subcategory of coronary heart disease population, sample size, characteristics of home-based cardiac telerehabilitation program, approach, and timing of technology acceptance evaluation data were extracted by the first author and confirmed by the second author, who made adjustments and included additional information where necessary. Features of the home-based cardiac telerehabilitation programs were categorized according to recommendations by Whitelaw et al [16] to facilitate uptake of digital health interventions. We categorized the core components present in the home-based cardiac telerehabilitation programs using American Heart Association classifications [6].

Collating, Summarizing, and Reporting the Findings

We used a 3-phase process [17] to systematically conduct our content analysis of the technology acceptance of home-based cardiac telerehabilitation among patients with coronary heart disease: preparing, organizing, and reporting of data. Because the aim of the review was to structure a descriptive analysis of home-based cardiac telerehabilitation acceptance based on the constructs of the technology acceptance model, we used a deductive content analysis approach [18,19].

A structured categorization matrix was prepared based on the constructs of the technology acceptance model: usability, utility, acceptability, acceptance, and external variables. Two authors concurrently and independently reviewed all the studies for content and coded data that corresponded to categories in the matrix. Content that did not fit into the other categories was gathered and coded under the category *external variables* and was analyzed based on the principle of inductive content analysis—open coding was undertaken, and data were grouped into similar categories and labeled with subcategories using content-characteristic words [17]. After data were organized, each author reviewed all of the studies under each category to check the reliability of the content analysis process and identify discrepancies in the collating and categorization of study data. Discussions were held until both authors were in agreement with the content under each category. Consultation with a third author was used to resolve any disagreements. The timing of home-based cardiac telerehabilitation evaluations was categorized based on when evaluation was undertaken relative to the trial implementation stage—pretrial, intratrial, or posttrial.

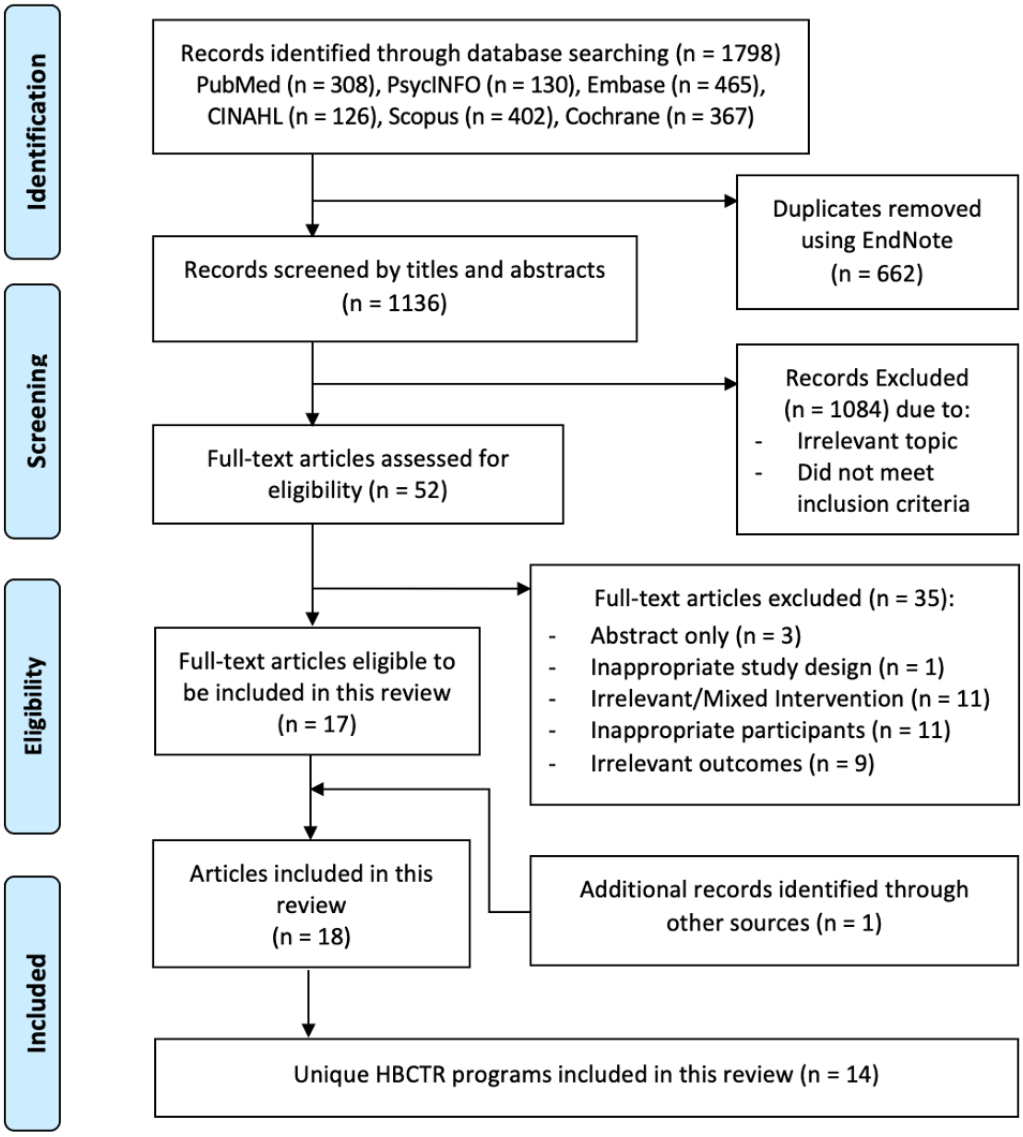
Quality appraisal was not performed as the objective of this scoping review was to provide an overview of the existing evidence on the evaluations of usability, utility, acceptability, and acceptance in home-based cardiac telerehabilitation, regardless of the quality of the evidence [12].

Results

General

The search generated 1136 unique papers. After title and abstract screening, 1084 papers were excluded. The remaining 52 full-text papers were retrieved and screened, and 35 papers were excluded (Table S2 in Multimedia Appendix 1). Manual searches of the reference lists of relevant papers identified 1 paper for inclusion; therefore, 18 papers [20-37], with 14 independent home-based cardiac telerehabilitation programs, were included in this review (Figure 1).

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) flow diagram. HBCTR: home-based cardiac telerehabilitation.



Characteristics of Studies

Studies included in this review (Table S3 in Multimedia Appendix 2) were published between 2007 and 2021; the majority (n=14) were published after 2013. Studies were conducted in the following countries: China [23,24,28,31,33,34]; Australia [29,30,32,36]; Canada [35-37]; United States of America [22,25]; United Kingdom [20,21]; and New Zealand [27]. Studies included patients who had the following: stable angina; myocardial infarction; stable coronary heart disease; or underwent coronary revascularization (ie, coronary artery bypass grafting or percutaneous coronary intervention). In studies that reported age and gender of participants, the mean age of patients ranged from 53 to 66 years and the proportion of female patients ranged from 9.4% to 33%. Devi et al [21] and Varnfield et al [29] were earlier papers reporting on the same home-based cardiac telerehabilitation programs as those in Devi et al [20] and Varnfield et al [30], respectively. Zutz et al [35] and Lear et al [36] were both earlier papers reporting on the same home-based cardiac telerehabilitation program as that in Banner et al [37].

Characteristics of Home-Based Cardiac Telerehabilitation Programs

Home-based cardiac telerehabilitation programs were delivered mainly via smartphone apps (n=11) and websites (n=3) and were supplemented by other modes of delivery: text messaging (n=6), telephone calls (n=5), emails (n=2), videoconferencing (n=1), and telemonitoring (n=10). Telemonitoring devices that supported remote supervision of exercise training by the cardiac rehabilitation team and patients' self-monitoring of physical activity included heart rate monitors, accelerometers, and pedometers.

Features of the home-based cardiac telerehabilitation programs included engagement of stakeholders, clinicians, and patients throughout the design or development of the home-based cardiac telerehabilitation program (n=3); testing of the home-based cardiac telerehabilitation program by cardiology experts and patients (n=8); provision of face-to-face training on use of home-based cardiac telerehabilitation for patients (n=10); ongoing technical support throughout home-based cardiac telerehabilitation program (n=4); and consideration of data

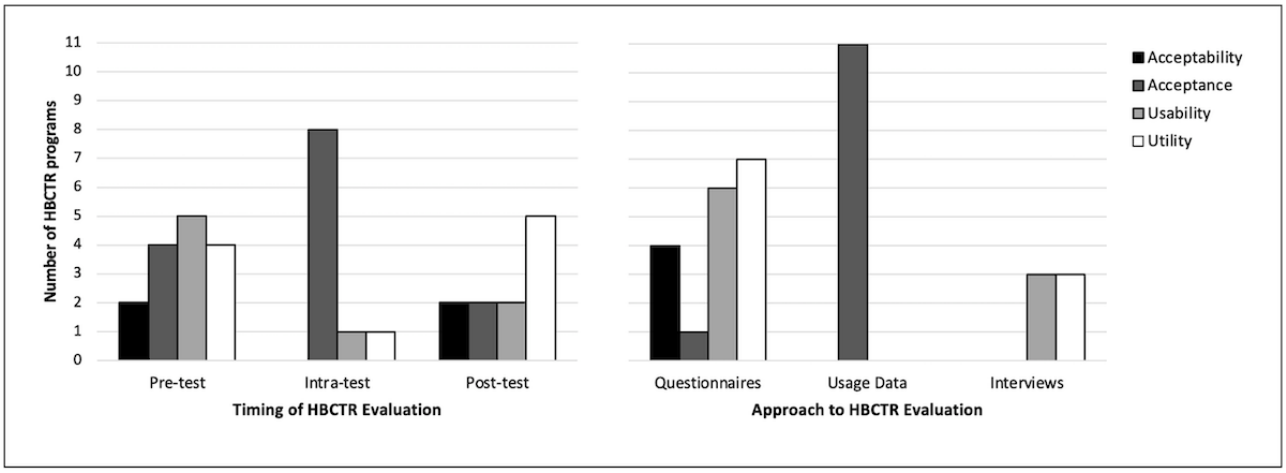
privacy and security in the use of technologies in home-based cardiac telerehabilitation (n=7).

The American Heart Association core components [6] that were present in the home-based cardiac telerehabilitation programs were patient assessment (n=14), exercise training (n=13), dietary management (n=10), risk factor management (n=11), medication adherence (n=8), and psychosocial support (n=6). Only 5 studies [23,24,26,30,31] had a comprehensive home-based cardiac telerehabilitation program that included all the core components (Table S4 in Multimedia Appendix 2).

Timing and Approaches to Evaluation

Home-based cardiac telerehabilitation programs were commonly evaluated at the pretrial stage (n=5) and using a combination of intra and posttrial measures (n=4), followed by intratrial only (n=3), posttrial only (n=1) and a combination of pre-, intra-, and posttrial measures (n=1) (Figure 2). The following methods were used to evaluate home-based cardiac telerehabilitation: questionnaires (n=10); usage data (n=11); and interviews (n=3). Except for one study [22] that used the System Usability Scale, the remaining questionnaires used were ad hoc surveys. Yu et al [33] used a combination of captured usage data and patient-report questionnaires to evaluate acceptance of home-based cardiac telerehabilitation at both the intra- and posttrial stage. Higgins et al [26] used both questionnaires and interviews to evaluate both usability and utility of their home-based cardiac telerehabilitation program.

Figure 2. Evaluation timing (left) and approach (right) over the dimensions of the technology acceptance model constructs. HBCTR: home-based cardiac telerehabilitation.



Usability

Of the 18 studies, 7 studies [22,25-27,29,32,37] reported the usability of home-based cardiac telerehabilitation programs. Specific outcomes measures within the usability construct included perceived ease of system use and navigation [22,25,29,32], ease and comfort of use of wearable devices [27], system learnability [22,26], and comprehension and ease of undertaking tasks on the system [27,37]. Overall, studies reported high usability rating scores and qualitative feedback from participants regarding home-based cardiac telerehabilitation use.

Utility

Specific outcomes measures within the utility construct included perceived usefulness in supporting behavior change [21,23,26,27,37], in managing psychological well-being [21,26], in controlling symptoms [21], in tracking goals and progress [21,25-27,37], in reducing outpatient visits [23], and of the overall home-based cardiac telerehabilitation system [23,29,33]. Utility of home-based cardiac telerehabilitation was generally favorably perceived, with the exception of 2 studies [29,33] in which perceived usefulness of the system was rated poorly.

Acceptability

High rates of acceptability were reported in 3 studies [22,27,33], ranging from 81.3% to 88% of participants who agreed that they would continue to use the home-based cardiac telerehabilitation system regularly after they had completed the study intervention period. Prior to system use, one study [24] reported an acceptability rate of 59.3% (participants who were potentially willing to participate in a home-based cardiac telerehabilitation program).

Acceptance

Most studies reported participants' usage of the home-based cardiac telerehabilitation system either through direct evaluation of program usage data or through self-reported participant survey responses (Table 1). Studies included a very broad range of outcome measures including engagement with home-based cardiac telerehabilitation [20,22,23,33,35,36] (ie, frequency and volume of website log-ins, smartphone app usage, activity tracker wear time); tasks completed [22,23,25,31,33] (ie, frequency and volume of educational modules reviewed, vitals logged, counseling sessions attended, response to program reminders), and captured exercise data [22,27,28,30,32,34] (ie, objective telemonitoring data on the uptake, adherence, and completion of prescribed exercise sessions and goals). Overall, usage was high, reflecting high end-user acceptance. Only 5 studies [22,25,30,33,34] reported usage data for specific components over time to determine the timepoints when participant usage tapered or ceased (Figure 3).

Table 1. Acceptance of home-based cardiac telerehabilitation programs.

Method, definition of actual use, and data timepoint	Acceptance of home-based cardiac telerehabilitation program
Program usage data	
Engagement with home-based cardiac telerehabilitation program	
6 weeks	• Mean total number of 29 website log-ins (range 7-44; average 5 times per week) [20,21]
12 weeks	• Mean total number of 50 website log-ins (range 26-86; average 4.2 times per week) [35] • Wearable worn for a median of 61 of 84 study days (IQR 35-78) for a median of 12.7 hours (IQR 11.1-13.8) per day [22]; • Mean decrease in wear time of 0.06 hours per week over 12 weeks [22]
16 weeks	• Mean total number of 27 website log-ins (range 0-140) [36]
24 weeks	• Proportion of participants who used and operated the app was 88.1% (4 weeks); 42.5% (8 weeks); 26.3% (12 weeks); 13.0% (16 weeks); 10.2% (20 weeks); 9.2% (24 weeks) [33]
Tasks completed	
12 weeks	• Participants completed an average off 66% (range 12.5%-100%) of weekly tasks (ie, intake form, heart rate upload, blood pressure data entry) [35] • Median number of 11 weekly telephone counseling sessions attended; 91.7% of weekly telephone counseling sessions completed [22] • Blood pressure recordings logged 3.6 (SD 2.1) times per week (at 4 weeks) and 3.6 (SD 1.9) (at 12 weeks); weight recordings logged 3.3 (SD 2.2) times per week (at 4 weeks) and 3.4 (SD 1.7) (at 12 weeks); mean 26.3 (SD 17.2) health-related messages text messages sent; reported exercises that met prespecified target heart rate an average of 3.5 (SD 1.4) times per week (at 4 weeks) and 3.5 (SD 1.1) times (at 12 weeks) [25]
16 weeks	• 41% of participants uploaded ≥32 exercise reports (average 2 exercise sessions per week); 26% of participants uploaded the required 8 blood pressure reports throughout study [36] • Total of 122 individual chat sessions (mean 3.6 per participant) with either nurse, dietician, or exercise specialist [36] • Participants used an average of 2.4, 2.6, and 2.7 hours of nursing, dietitian, and exercise specialist time, respectively [36]
24 weeks	• Proportion of participants who responded to medication reminders and health questionnaires was 34% (4 weeks); 21.2% (8 weeks); 14.2% (12 weeks); 11% (16 weeks); 8.3% (20 weeks); 7.7% (24 weeks) [33]
52 weeks	• 96.3% of participants read education papers 4 times per month; 98.8% of participants consulted with their health care managers 1-4 times per month; 82.7% of participants sent their test results (ie, blood pressure and blood results) 4-8 times over 52 weeks [31]
Captured exercise data	
6 weeks	• 86.6% of participants completed scheduled exercise sessions [32] • Uptake[a] rate: 80%; adherence[b] rate: 94%; completion[c] rate: 80% [30]
8 weeks	• Uptake rate: 87%; adherence rate: 75%; completion[d] rate: 75% [34]
12 weeks	• 86% of prescribed exercise goals completed over the 12-week study period; average decline of 8% completion per additional study week; 34% of walking goals completed over the 12-week study period; mean weekly increase in completion rate of 1% per additional week [22] • Adherence rate to prescribed exercise was 58.34% (range 0-100) [27]
24 weeks	• Participants exercised an average of 5.1 (SD 0.6) times a week; each time was 31.4 (SD 4.5) minutes [28]

Method, definition of actual use, and data timepoint	Acceptance of home-based cardiac telerehabilitation program
Self-reported survey responses	
Engagement with home-based cardiac telerehabilitation program	
24 weeks	• 100% of participants received WeChat modules and messages [23] • 17.4% of participants reported using the app every day; 44.6% of participants often forgot to use the app [33]
Tasks completed	
24 weeks	• 95% of participants read 75%-100% of WeChat modules and messages; 89% of participants read WeChat modules more than twice) [23]

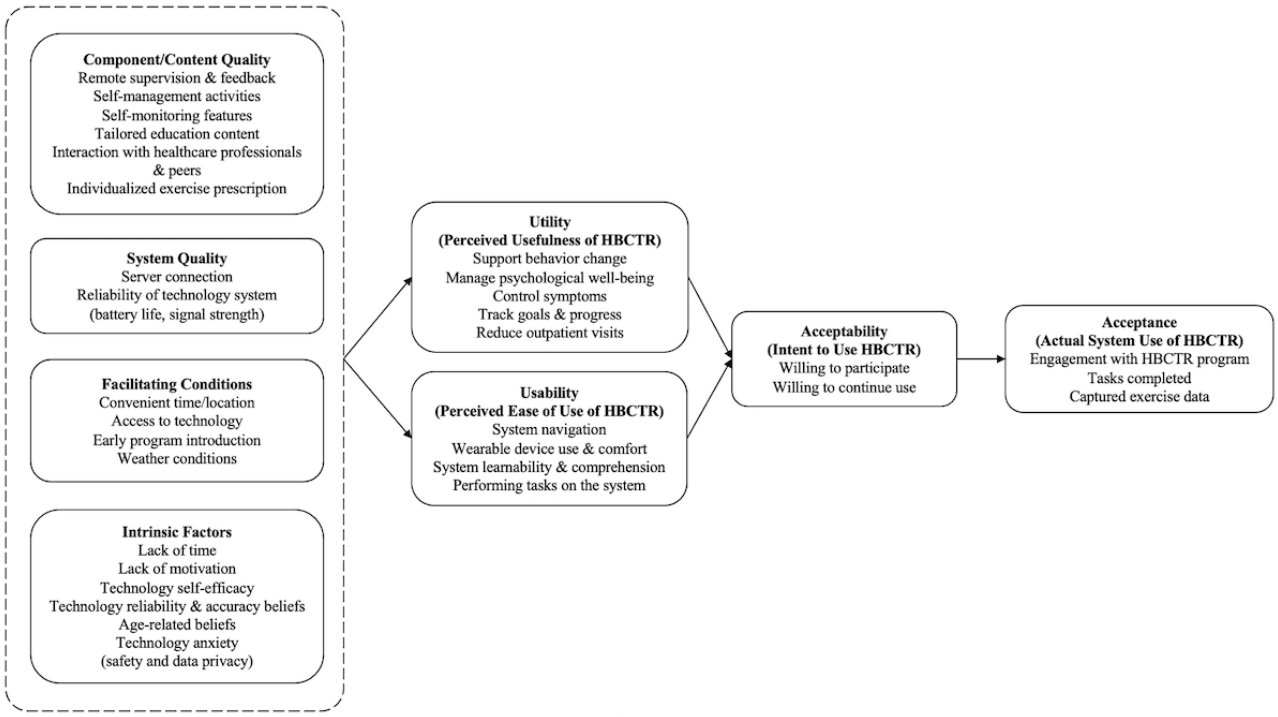
[a]Uptake was defined as attending baseline assessment, and uploading exercise data once to the home-based cardiac telerehabilitation platform.

[b]Adherence was defined as uploading 4 weeks of exercise data onto the home-based cardiac telerehabilitation.

[c]Completion was defined as attendance at the 6-week assessment.

[d]Completion was defined as attendance at the 8-week assessment.

Figure 3. Technology acceptance of home-based cardiac telerehabilitation (HBCTR) programs.



External Variables

Component Quality

The majority of the existing literature (n=8) on home-based cardiac telerehabilitation evaluation reported the program components that participants valued: remote supervision and feedback [21,24,27-29,37], support for self-management and self-monitoring [21,24,27,35,37], range of relevant educational modules [21,26,28], ability to communicate with health care professionals [21,27,35], and individualized exercise prescription [27]. Participants desired more interactive components such as chat platforms and noticeboards with peers to facilitate peer interaction and support [26,27,35] and greater intra- and postprogram support [27]. Participants in one study [26] wanted specific education content pertaining to death anxiety, and content that aligned rehabilitation goals with the purpose of living.

System Quality

Two studies [32,35] detailed participants' perspectives on the technical efficiency of the home-based cardiac telerehabilitation system. Specifically, issues relating to server connection and reliability of the technology (ie, equipment battery life and signal strength) were reported as these influenced participants ability to engage with the program without interruption.

Intrinsic Factors

Participants reported several intrinsic factors at the individual level that influenced how they perceived home-based cardiac telerehabilitation programs (n=5). These included lack of time [21,27,37], lack of motivation [21,27,37], perceived self-efficacy in operating the telerehabilitation system [28], perceived reliability and accuracy of technology [24], apprehension related to safety and data privacy [24], and preconceived beliefs

regarding the suitability of home-based cardiac telerehabilitation for older age [21,27].

Facilitating Conditions

The existence of resource and situational factors facilitated the usage of home-based cardiac telerehabilitation programs in included studies (n=5). Participants valued the accessibility and convenience offered by home-based cardiac telerehabilitation as it overcame restrictions related to time and location [21,27,37], but some expressed that regular access to the internet and computers would have facilitated uninterrupted usage of the program in earlier studies [29,37]. Situational factors such as timing of program introduction also influenced participants' perception of home-based cardiac telerehabilitation usage [21,26]. Participants reported wanting the program to begin sooner after their diagnosis to facilitate early establishing of routines and prevent potential cardiac complications [21,26]. Wet and cold seasons were reported as a barrier to outdoor physical exercises [21].

Details of the external variables, usability, utility, and acceptability of home-based cardiac telerehabilitation reported in included studies can be found in Table S5 (Multimedia Appendix 2).

Discussion

Principal Findings

In our scoping review, we found that most evaluations were undertaken at the intratrial and posttrial stage using singular methodological approaches, and although home-based cardiac telerehabilitation had high usability, utility, acceptability, and acceptance, patients reported a number of external variables such as component quality, system quality, intrinsic factors, and facilitating conditions that influenced how they interacted with the home-based cardiac telerehabilitation program.

Timing of Home-Based Cardiac Telerehabilitation Evaluation

Early evaluation of end-user acceptance and feasibility issues can critically inform the development and design of digital interventions and mitigate risks that an intervention is later undesirable or even abandoned at trial implementation stages [10,38]. Through our scoping review, we found that the majority of home-based cardiac telerehabilitation programs reported evaluations of technology acceptance either during or after trial implementation; evaluations were rarely reported at the pretrial stage. This may reflect a tendency in implementation research to prioritize the evaluation of trial intervention effectiveness over trial implementation effectiveness [39]. Yet, achieving intended trial effects is greatly dependent on participants' sufficient engagement with the implemented technology in a trial that strongly appeals to their contextual health care needs [10]. Hence, there is a need for future research on home-based cardiac telerehabilitation to refocus efforts of program evaluation more upstream, so that identified technology acceptance issues can be addressed and programs finetuned to ensure optimal success before trial implementation.

Approaches to Home-Based Cardiac Telerehabilitation Evaluation

Our review of the methodological approaches used to evaluate the technology acceptance constructs in home-based cardiac telerehabilitation revealed 3 main concerns. First, although home-based cardiac telerehabilitation programs used either quantitative (ie, survey questionnaires) or qualitative (ie, interviews) approaches to evaluate usability, utility, and acceptability, only 3 studies [21,26,37] employed qualitative methods (Figure 2), and only one study [26] used both approaches in tandem to evaluate the same technology acceptance attribute. Questionnaires are usually inexpensive and useful in gathering quantitative data in large samples but lack the ability to facilitate comprehension of in-depth individual variation in behaviors, perspectives, and experiences that qualitative interviews provide [40]. Such information is crucial to designing and delivering home-based cardiac telerehabilitation programs that truly match patients' needs and preferences. We recommend that future home-based cardiac telerehabilitation programs employ a mixed methods approach, comprising both quantitative and qualitative methods to guarantee evaluation results that are practical, interpretable, and comprehensive [40].

Second, apart from one study [22] that used the System Usability Scale questionnaire, the remaining home-based cardiac telerehabilitation programs in this review used customized ad hoc questionnaires to measure the constructs of technology acceptance. This corresponds with the findings of previous reviews [38,41], which mostly included studies that evaluated digital health technology acceptance attributes using quantitative measures that lacked the psychometric properties of reliability and validity. This finding highlights an apparent scarcity of validated tools to evaluate technology acceptance in the context of digital health [38]. Furthermore, this could reflect the need for researchers to develop their own questionnaires that consider program-specific components, with general acceptance concepts, to allow for an assessment of technology acceptance attributes that is tailored to the particular home-based cardiac telerehabilitation context and population. However, this makes comparing results across studies challenging. It would be commendable to see future research efforts dedicated to adapting existing questionnaires or even validating new tools that encompass the unique home-based cardiac telerehabilitation context. We believe that having such generalizable measures can greatly advance home-based cardiac telerehabilitation research and practice by creating opportunities for comparable data on technology acceptance constructs to be analyzed and for comparative benchmarks to be set in home-based cardiac telerehabilitation program evaluation.

Third, home-based cardiac telerehabilitation programs had varied definitions and measurements of acceptance (ie, actual system usage) (Table 1). This is consistent with previous literature on the use of digital health technologies for cardiovascular disease self-management [42] and may be indicative of attempts to examine the multifarious behavior changes addressed in cardiac rehabilitation. Given that user engagement with technology is a dynamic process occurring in a self-directed manner by which users continually decide to

either use or abandon a technology system [38,43], evaluations of home-based cardiac telerehabilitation acceptance should account for this temporal nature and analyze how usage evolves over the course of the rehabilitation program. This is especially important as interventions such as home-based cardiac telerehabilitation are theorized to require sustained use over time to realize intended effects. However, only 5 home-based cardiac telerehabilitation programs [22,25,30,33,34] reported usage over time (date-tagged acceptance data). Gallagher and Zhang [10] recommend the clear identification of individual digital health components targeted at behavior change and the integration of software capabilities that can monitor the usage of respective components. As the eventual goal of home-based cardiac telerehabilitation programs is successful incorporation into clinical practice, it would be interesting to see future studies examine the causal relationships between the level of home-based cardiac telerehabilitation usage and objective intervention outcome over time to determine the specific dose of a home-based cardiac telerehabilitation component needed to achieve optimal behavioral, physiological, and clinical outcomes.

Technology Acceptance of Home-Based Cardiac Telerehabilitation

The acceptance rates observed in our review could be explained by the high usability, utility, and acceptability reported in the programs and correspond to the fundamental basis of the technology acceptance model, that is, that technology acceptance is determined by the degree of value and perceived burden [11]. This finding not only offers validation to the technology acceptance model but points to the potential of home-based cardiac telerehabilitation to revolutionize the landscape of secondary prevention by blending traditional services provided by health care professionals with technology-enabled self-care platforms to continue the provision of patient-centered care. This is especially crucial during the current COVID-19 pandemic to mitigate the demand for in-person services [4]. The suitability of home-based cardiac telerehabilitation as an effective alternative to center-based cardiac rehabilitation has been recently reported [7], with prospects for significant economic cost-savings through improved productivity and health outcomes [44]. Yet, an evaluation of end-user acceptance is foundational if barriers and gaps to patient uptake are to be addressed, and if successful wide-scale implementation of home-based cardiac telerehabilitation into clinical practice is to be realized. In the context of home-based cardiac telerehabilitation for patients with coronary heart disease, our review underlined the external variables that have influenced patient's perceived usability and utility of home-based cardiac telerehabilitation. Recommendations for addressing these variables are offered in in the following paragraphs and may serve to provide a foundation for the development and design of future home-based cardiac telerehabilitation programs (Table 2).

Table 2. Recommendations to improve home-based cardiac telerehabilitation acceptance and its evaluation.

Topic	Recommendation
Evaluation timing	Home-based cardiac telerehabilitation program evaluation should be undertaken throughout the entirety of the developmental and implementation, ie, before, during and after trial implementation.
Evaluation approach	Home-based cardiac telerehabilitation program evaluation should employ a mixed approach comprising of both quantitative and qualitative methods.
	Measurement tools must be tailored to encompass the unique context of home-based cardiac telerehabilitation by adapting existing questionnaires or validating new ones.
	Evaluations of home-based cardiac telerehabilitation technology acceptance should analyze how usage of individual program components evolves over the course of the rehabilitation program.
	Causal relationships between home-based cardiac telerehabilitation usage and intervention outcomes should be examined to determine specific doses needed to achieve optimal behavioral, physiological, and clinical outcomes.
Design and testing	Developers should prioritize user-centered approaches by partnering with end users (ie, clinicians and patients) in the co-designing of programs in the early stages of program design.
	Field-testing and evaluations of the technologies supporting home-based cardiac telerehabilitation services should occur prior to trial implementation stages.
Individualization	Home-based cardiac telerehabilitation programs should be offered as early as possible for patients.
	Alternatives for either indoor or outdoor exercise training should be programmed.
Accessibility	Home-based cardiac telerehabilitation programs should be adapted to the socioeconomic needs of end users and their community
	Partnerships with local governing bodies should be established to marshal resources and secure funding to invest in required infrastructure.
	The prospects of insurance coverage for home-based cardiac telerehabilitation programs should be explored.
	home-based cardiac telerehabilitation programs should be reasonably priced with subsidies for mobile phones, data plans and wearables.
Data privacy and security	Home-based cardiac telerehabilitation should provide patients with transparent privacy policies and comply with data governance regulations and security protocols.
Training	Patients should be provided introductory training sessions that are supported by practical step-by-step instruction manuals.
Technology support	Designated technical support staff should be made available on home-based cardiac telerehabilitation platforms.

Recommendations for Home-Based Cardiac Telerehabilitation Development

Home-based cardiac telerehabilitation developers should prioritize user-centered approaches by partnering with end users (ie, clinicians and patients) in the co-design, field test, and evaluation of technologies supporting telerehabilitation services [10]. Accounting for the needs and preferences of patients in the early stages of program design can help mitigate concerns regarding home-based cardiac telerehabilitation component quality and can help in identifying issues with home-based cardiac telerehabilitation system quality program testing prior to trial implementation stages. However, we observed that less than one-fifth of home-based cardiac telerehabilitation programs reported including end users in the design and development stage and just over half undertook user testing of the home-based cardiac telerehabilitation system (Table S4 in Multimedia Appendix 2). American Heart Association's recommendations on home-based cardiac telerehabilitation [6] and the Beatty et al [45] framework for mobile technology in cardiac rehabilitation can guide the development and evaluation of future home-based cardiac telerehabilitation programs.

Facilitating conditions, such as the timing of program introduction, prevailing weather conditions, and access to internet and computers, were reported to influence patients' use of home-based cardiac telerehabilitation programs. Given that peak lifestyle changes occur in the first 6 months after diagnosis [46] and that early cardiac rehabilitation is a significant predictor of cardiac function and functional capacity [47,48], home-based cardiac telerehabilitation should be offered as early as possible for patients to ensure optimal outcomes. Home-based cardiac telerehabilitation should also offer patients alternatives for either indoor or outdoor exercise training, especially in regions with seasonal weather changes. Additionally, as inequities in cardiovascular health still exist, an examination of socioeconomic characteristics are crucial if technology accessibility and affordability issues surrounding home-based cardiac telerehabilitation usage are to be addressed [49,50]. Access to technology infrastructure remains unevenly distributed worldwide, with internet use being significantly lower in low- and middle-income regions than in high-income regions [51]. Partnerships with local governing bodies should be established to marshal resources and secure funding to invest in required infrastructure [52]. Collaborating with nongovernment organizations to advocate for prospects on insurance coverage and to negotiate reasonable pricing and subsidies for mobile phones, data plans, and wearables will aid in supporting the long-term implementation and scale-up of home-based cardiac telerehabilitation in clinical practice [53].

Although intrinsic factors such as lack of time and motivation are less amenable to change, program adaptations can be made to palliate concerns regarding data privacy, perceived technology self-efficacy and reliability, and preconceived age-related beliefs regarding home-based cardiac telerehabilitation usage. Program training, technological support, and the availability of transparent privacy policies, especially for older adults, can reduce potential uneasiness and facilitate willingness to engage in digital health technologies such as home-based cardiac telerehabilitation [4,54,55]. Even though the majority of included home-based cardiac telerehabilitation provided face-to-face program training, less than one-third offered ongoing technological support during program intervention, and only half indicated using secure password-protected platforms (Table S4 in Multimedia Appendix 2). Future programs should develop introductory training sessions that are supported by practical step-by-step instruction manuals with designated technical support staff on home-based cardiac telerehabilitation platforms that comply with data governance regulations and security protocols to mitigate the risk of privacy breaches [54,55].

Limitations

This scoping review has some limitations that need to be acknowledged. First, the inclusion of only English-language papers may have resulted in the omission of eligible papers published in other languages. However, our comprehensive search strategy and broad inclusion of different study designs with no time restrictions allows for breadth and depth of inclusion in this review. Second, although the technology acceptance model offers a user-centered approach in mapping patient perspectives of home-based cardiac telerehabilitation program acceptance, content analysis is inherently reductive and could have limited the scope of our findings. However, the thematic analysis undertaken to explore the external variables influencing home-based cardiac telerehabilitation acceptance could have mitigated the risks of missing meaningful data from the studies included in our review. Lastly, although end users in this user-centered approach also include health care providers delivering home-based cardiac telerehabilitation, the evaluation of technology acceptance from provider perspectives was not included because it was not the focus of this review. It is likely that the underlying determinants of home-based cardiac telerehabilitation acceptance may differ in these users. We recommend that future research in the field of home-based cardiac telerehabilitation aim to include literature in other languages, utilize other available conceptual frameworks on digital health acceptance, and accommodate perspectives from different categories of end users in order to fully comprehend and address home-based cardiac telerehabilitation implementation and acceptance.

Conclusions

We drew on the technology acceptance model to map available research on patient's technology acceptance of home-based cardiac telerehabilitation. Our results demonstrated that, while patient perspectives on home-based cardiac telerehabilitation usability, utility, acceptability, and acceptance were high, a number of external variables influence technology acceptance of home-based cardiac telerehabilitation programs. Additionally, gaps in current home-based cardiac telerehabilitation evaluation timing and approaches were revealed. As the appeal for home-based cardiac telerehabilitation grows during the COVID-19 pandemic and beyond, findings from this review can be used to provide guidance for stakeholders and clinicians in developing and evaluating patient-centered home-based cardiac telerehabilitation programs.

Acknowledgments

The assistance of Ms. Annelissa Chin with the peer-review database search strategy is greatly appreciated. This project received funding (grant MOH-000364) from the National Medical Research Council, under the Ministry of Health Singapore.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 131 KB - jmir_v24i1e34657_app1.pdf]

Multimedia Appendix 2
Summary, characteristics, and outcomes of studies.
[PDF File (Adobe PDF File), 394 KB - jmir_v24i1e34657_app2.pdf]

References

1. Mensah GA, Roth GA, Fuster V. The global burden of cardiovascular diseases and risk factors: 2020 and beyond. J Am Coll Cardiol 2019 Nov 19;74(20):2529-2532 [FREE Full text] [doi: 10.1016/j.jacc.2019.10.009] [Medline: 31727292]
2. Ambrosetti M, Abreu A, Corrà U, Davos CH, Hansen D, Frederix I, et al. Secondary prevention through comprehensive cardiovascular rehabilitation: from knowledge to implementation. 2020 update. a position paper from the secondary

prevention and rehabilitation section of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2020 Apr 07:2047487320913379. [doi: 10.1177/2047487320913379] [Medline: 33611446]
3. Santiago de Araújo Pio C, Chaves G, Davies P, Taylor R, Grace S. Interventions to promote patient utilisation of cardiac rehabilitation. Cochrane Database Syst Rev 2019 Feb 01;2:CD007131 [FREE Full text] [doi: 10.1002/14651858.CD007131.pub4] [Medline: 30706942]
4. Scherrenberg M, Wilhelm M, Hansen D, Völler H, Cornelissen V, Frederix I, et al. The future is now: a call for action for cardiac telerehabilitation in the COVID-19 pandemic from the secondary prevention and rehabilitation section of the European Association of Preventive Cardiology. Eur J Prev Cardiol 2020 Jul 02:2047487320939671. [doi: 10.1177/2047487320939671] [Medline: 32615796]
5. Ghisi GLDM, Xu Z, Liu X, Mola A, Gallagher R, Babu AS, et al. Impacts of the COVID-19 pandemic on cardiac rehabilitation delivery around the world. Glob Heart 2021 Jun 10;16(1):43-43 [FREE Full text] [doi: 10.5334/gh.939] [Medline: 34211829]
6. Thomas RJ, Beatty AL, Beckie TM, Brewer LC, Brown TM, Forman DE, et al. Home-based cardiac rehabilitation: a scientific statement from the American Association of Cardiovascular and Pulmonary Rehabilitation, the American Heart Association, and the American College of Cardiology. Circulation 2019 Jul 02;140(1):e69-e89 [FREE Full text] [doi: 10.1161/CIR.0000000000000663] [Medline: 31082266]
7. Ramachandran H, Jiang Y, Tam W, Yeo T, Wang W. Effectiveness of home-based cardiac telerehabilitation as an alternative to phase 2 cardiac rehabilitation of coronary heart disease: a systematic review and meta-analysis. Eur J Prev Cardiol 2021 Jul 13 [FREE Full text] [doi: 10.1093/eurjpc/zwab106] [Medline: 34254118]
8. Frederix I, Caiani EG, Dendale P, Anker S, Bax J, Böhm A, et al. ESC e-Cardiology working group position paper: overcoming challenges in digital health implementation in cardiovascular medicine. Eur J Prev Cardiol 2019 Jul;26(11):1166-1177. [doi: 10.1177/2047487319832394] [Medline: 30917695]
9. Saarijärvi M, Wallin L, Bratt E. Process evaluation of complex cardiovascular interventions: how to interpret the results of my trial? Eur J Cardiovasc Nurs 2020 Mar;19(3):269-274 [FREE Full text] [doi: 10.1177/1474515120906561] [Medline: 32054300]
10. Gallagher R, Zhang L. Evaluating mobile health technologies: does the traditional randomized controlled trial serve our needs? Eur J Cardiovasc Nurs 2021 Aug 20;20(6):623-626. [doi: 10.1093/eurjcn/zvab053] [Medline: 34269387]
11. Davis FD, Venkatesh V. A critical assessment of potential measurement biases in the technology acceptance model: three experiments. Int J Hum Comput Stud 1996 Jul;45(1):19-45. [doi: 10.1006/ijhc.1996.0040]
12. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb 23;8(1):19-32. [doi: 10.1080/1364557032000119616]
13. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E. Systematic review or scoping review? guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 2018 Dec 19;18(1):143-143 [FREE Full text] [doi: 10.1186/s12874-018-0611-x] [Medline: 30453902]
14. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
15. Nielsen J. Usability Engineering. San Francisco: Morgan Kaufmann; 1994.
16. Whitelaw S, Pellegrini D, Mamas M, Cowie M, Van Spall HGC. Barriers and facilitators of the uptake of digital health technology in cardiovascular care: a systematic scoping review. Eur Heart J Digit Health 2021 Mar;2(1):62-74 [FREE Full text] [doi: 10.1093/ehjdh/ztab005] [Medline: 34048508]
17. Elo S, Kyngäs H. The qualitative content analysis process. J Adv Nurs 2008 Mar 18;62(1):107-115. [doi: 10.1111/j.1365-2648.2007.04569.x] [Medline: 18352969]
18. Sandelowski M. Qualitative analysis: what it is and how to begin. Res Nurs Health 1995 Aug 01;18(4):371-375. [doi: 10.1002/nur.4770180411] [Medline: 7624531]
19. Kyngas H, Vanhanen L. Content analysis as a research method [Finnish]. Hoitotiede 1999;11:3-12.
20. Devi R, Powell J, Singh S. A web-based program improves physical activity outcomes in a primary care angina population: randomized controlled trial. J Med Internet Res 2014 Sep 12;16(9):e186 [FREE Full text] [doi: 10.2196/jmir.3340] [Medline: 25217464]
21. Devi R, Carpenter C, Powell J, Singh S. Exploring the experience of using a web-based cardiac rehabilitation programme in a primary care angina population: a qualitative study. Int J Ther Rehabil 2014 Sep 02;21(9):434-440. [doi: 10.12968/ijtr.2014.21.9.434]
22. Ding EY, Erskine N, Stut W, McManus DD, Peterson A, Wang Z, et al. MI-PACE home-based cardiac telerehabilitation program for heart attack survivors: usability study. JMIR Hum Factors 2021 Jul 08;8(3):e18130 [FREE Full text] [doi: 10.2196/18130] [Medline: 34255660]
23. Dorje T, Zhao G, Tso K, Wang J, Chen Y, Tsokey L, et al. Smartphone and social media-based cardiac rehabilitation and secondary prevention in China (SMART-CR/SP): a parallel-group, single-blind, randomised controlled trial. Lancet Digit Health 2019 Nov;1(7):e363-e374 [FREE Full text] [doi: 10.1016/S2589-7500(19)30151-7] [Medline: 33323210]

24. Fang J, Li J, Li ZH, Xu DM, Chen C, Xie B, et al. Attitudes towards acceptance of an innovative home-based and remote sensing rehabilitation protocol among cardiovascular patients in Shantou, China. J Geriatr Cardiol 2016 May;13(4):326-332 [FREE Full text] [doi: 10.11909/j.issn.1671-5411.2016.04.006] [Medline: 27403142]
25. Harzand A, Witbrodt B, Davis-Watts ML, Alrohaibani A, Goese D, Wenger NK, et al. Feasibility of a smartphone-enabled cardiac rehabilitation program in male veterans with previous clinical evidence of coronary heart disease. Am J Cardiol 2018 Nov 01;122(9):1471-1476 [FREE Full text] [doi: 10.1016/j.amjcard.2018.07.028] [Medline: 30217377]
26. Higgins RO, Rogerson M, Murphy BM, Navaratnam H, Butler MV, Barker L, et al. Cardiac rehabilitation online pilot: extending reach of cardiac rehabilitation. J Cardiovasc Nurs 2017;32(1):7-13. [doi: 10.1097/JCN.0000000000000297] [Medline: 26422638]
27. Rawstorn JC, Gant N, Rolleston A, Whittaker R, Stewart R, Benatar J, et al. End users want alternative intervention delivery models: usability and acceptability of the Remote-CR exercise-based cardiac telerehabilitation program. Arch Phys Med Rehabil 2018 Nov 01;99(11):2373-2377. [doi: 10.1016/j.apmr.2018.06.027] [Medline: 30076800]
28. Song Y, Ren C, Liu P, Tao L, Zhao W, Gao W. Effect of smartphone-based telemonitored exercise rehabilitation among patients with coronary heart disease. J Cardiovasc Transl Res 2019 Dec 9;13(4):659-667 [FREE Full text] [doi: 10.1007/s12265-019-09938-6] [Medline: 31820334]
29. Varnfield M, Karunanithi MK, Särelä A, Garcia E, Fairfull A, Oldenburg BF, et al. Uptake of a technology-assisted home-care cardiac rehabilitation program. Med J Aust 2011 Feb 21;194(4):S15-S19. [doi: 10.5694/j.1326-5377.2011.tb02937.x] [Medline: 21401482]
30. Varnfield M, Karunanithi M, Lee C, Honeyman E, Arnold D, Ding H, et al. Smartphone-based home care model improved use of cardiac rehabilitation in postmyocardial infarction patients: results from a randomised controlled trial. Heart 2014 Nov;100(22):1770-1779 [FREE Full text] [doi: 10.1136/heartjnl-2014-305783] [Medline: 24973083]
31. Wang J, Zeng Z, Dong R, Sheng J, Lai Y, Yu J, et al. Efficacy of a WeChat based intervention to adherence to secondary prevention in patients undergoing coronary artery bypass graft in China: A randomized controlled trial. J Telemed Telecare 2020 Sep 30:1357633X20960639. [doi: 10.1177/1357633X20960639] [Medline: 32996349]
32. Worringham C, Rojek A, Stewart I. Development and feasibility of a smartphone, ECG and GPS based system for remotely monitoring exercise in cardiac rehabilitation. PLoS One 2011 Feb 09;6(2):e14669 [FREE Full text] [doi: 10.1371/journal.pone.0014669] [Medline: 21347403]
33. Yu C, Liu C, Du J, Liu H, Zhang H, Zhao Y, MISSION-2 Collaborative Group. Smartphone-based application to improve medication adherence in patients after surgical coronary revascularization. Am Heart J 2020 Oct;228:17-26. [doi: 10.1016/j.ahj.2020.06.019] [Medline: 32745732]
34. Yudi M, Clark D, Tsang D, Jelinek M, Kalten K, Joshi SB, et al. SMARTphone-based, early cardiac REHABilitation in patients with acute coronary syndromes: a randomized controlled trial. Coron Artery Dis 2021 Aug 01;32(5):432-440. [doi: 10.1097/MCA.0000000000000938] [Medline: 32868661]
35. Zutz A, Ignaszewski A, Bates J, Lear SA. Utilization of the internet to deliver cardiac rehabilitation at a distance: a pilot study. Telemed J E Health 2007 Jun;13(3):323-330. [doi: 10.1089/tmj.2006.0051] [Medline: 17603835]
36. Lear SA, Singer J, Banner-Lukaris D, Horvat D, Park JE, Bates J, et al. Randomized trial of a virtual cardiac rehabilitation program delivered at a distance via the internet. Circ Cardiovasc Qual Outcomes 2014 Nov;7(6):952-959. [doi: 10.1161/CIRCOUTCOMES.114.001230] [Medline: 25271050]
37. Banner D, Lear S, Kandola D, Singer J, Horvat D, Bates J, et al. The experiences of patients undertaking a 'virtual' cardiac rehabilitation program. Stud Health Technol Inform 2015;209:9-14. [Medline: 25980699]
38. Nadal C, Sas C, Doherty G. Technology acceptance in mobile health: scoping review of definitions, models, and measurement. J Med Internet Res 2020 Jul 06;22(7):e17256 [FREE Full text] [doi: 10.2196/17256] [Medline: 32628122]
39. Wensing M, Grol R. Knowledge translation in health: how implementation science could contribute more. BMC Med 2019 May 07;17(1):88 [FREE Full text] [doi: 10.1186/s12916-019-1322-9] [Medline: 31064388]
40. Creswell J, Creswell J. Research Design: Qualitative, Quantitative, and Mixed Methods Approaches Fifth ed. California: Sage Publications; 2018.
41. Sousa V, Dunn Lopez K. Towards usable e-health. a systematic review of usability questionnaires. Appl Clin Inform 2017 May 10;8(2):470-490 [FREE Full text] [doi: 10.4338/ACI-2016-10-R-0170] [Medline: 28487932]
42. Coorey G, Neubeck L, Mulley J, Redfern J. Effectiveness, acceptability and usefulness of mobile applications for cardiovascular disease self-management: systematic review with meta-synthesis of quantitative and qualitative data. Eur J Prev Cardiol 2018 Mar;25(5):505-521. [doi: 10.1177/2047487317750913] [Medline: 29313363]
43. Lee Y, Kozar KA, Larsen KR. The technology acceptance model: past, present, and future. Commun Assoc Inf Syst 2003;12(1):50. [doi: 10.17705/1cais.01250]
44. Jiang X, Ming W, You JH. The cost-effectiveness of digital health interventions on the management of cardiovascular diseases: systematic review. J Med Internet Res 2019 Jun 17;21(6):e13166 [FREE Full text] [doi: 10.2196/13166] [Medline: 31210136]
45. Beatty AL, Fukuoka Y, Whooley MA. Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation. J Am Heart Assoc 2013 Nov 01;2(6):e000568 [FREE Full text] [doi: 10.1161/JAHA.113.000568] [Medline: 24185949]

46. Chow CK, Jolly S, Rao-Melacini P, Fox KA, Anand SS, Yusuf S. Association of diet, exercise, and smoking modification with risk of early cardiovascular events after acute coronary syndromes. Circulation 2010 Feb 16;121(6):750-758. [doi: 10.1161/CIRCULATIONAHA.109.891523] [Medline: 20124123]
47. Prescott E, Eser P, Mikkelsen N, Holdgaard A, Marcin T, Wilhelm M, et al. Cardiac rehabilitation of elderly patients in eight rehabilitation units in western Europe: outcome data from the EU-CaRE multi-centre observational study. Eur J Prev Cardiol 2020 Nov;27(16):1716-1729. [doi: 10.1177/2047487320903869] [Medline: 32102550]
48. Giallauria F, Acampa W, Ricci F, Vitelli A, Maresca L, Mancini M, et al. Effects of exercise training started within 2 weeks after acute myocardial infarction on myocardial perfusion and left ventricular function: a gated SPECT imaging study. Eur J Prev Cardiol 2012 Dec;19(6):1410-1419. [doi: 10.1177/1741826711425427] [Medline: 21965517]
49. Oertelt-Prigione S, Maas AH. Health inequalities in secondary prevention. Eur J Prev Cardiol 2017 Jun;24(3_suppl):116-122. [doi: 10.1177/2047487317703767] [Medline: 28618901]
50. Niederseer D, Schmied C. Socioeconomic status matters: how can we individualize cardiac rehabilitation according to different socioeconomic needs? Eur J Prev Cardiol 2020 Jun 02:510-512. [doi: 10.1177/2047487320931309] [Medline: 32484046]
51. The state of mobile internet connectivity 2020. GSM Association. URL: https://www.gsma.com/r/wp-content/uploads/2020/09/GSMA-State-of-Mobile-Internet-Connectivity-Report-2020.pdf [accessed 2021-08-15]
52. Kruse C, Betancourt J, Ortiz S, Valdes Luna SM, Bamrah IK, Segovia N. Barriers to the use of mobile health in improving health outcomes in developing countries: systematic review. J Med Internet Res 2019 Oct 09;21(10):e13263 [FREE Full text] [doi: 10.2196/13263] [Medline: 31593543]
53. Liu P, Astudillo K, Velez D, Kelley L, Cobbs-Lomax D, Spatz ES. Use of mobile health applications in low-income populations: a prospective study of facilitators and barriers. Circ Cardiovasc Qual Outcomes 2020 Sep;13(9):e007031 [FREE Full text] [doi: 10.1161/CIRCOUTCOMES.120.007031] [Medline: 32885681]
54. Redfern J. Can older adults benefit from smart devices, wearables, and other digital health options to enhance cardiac rehabilitation? Clin Geriatr Med 2019 Nov;35(4):489-497. [doi: 10.1016/j.cger.2019.07.004] [Medline: 31543180]
55. Tsai HS, Rikard RV, Cotten SR, Shillair R. Senior technology exploration, learning, and acceptance (STELA) model: from exploration to use – a longitudinal randomized controlled trial. Educational Gerontology 2019 Nov 18;45(12):728-743. [doi: 10.1080/03601277.2019.1690802]

Edited by G Eysenbach; submitted 02.11.21; peer-reviewed by E van der Velde; comments to author 23.11.21; revised version received 24.11.21; accepted 27.11.21; published 07.01.22.

Please cite as:
Ramachandran HJ, Jiang Y, Teo JYC, Yeo TJ, Wang W
Technology Acceptance of Home-Based Cardiac Telerehabilitation Programs in Patients With Coronary Heart Disease: Systematic Scoping Review
J Med Internet Res 2022;24(1):e34657
URL: https://www.jmir.org/2022/1/e34657
doi:10.2196/34657
PMID:34994711

Review

Social Media and Health Care (Part II): Narrative Review of Social Media Use by Patients

Deema Farsi[1], BDS, MSc; Hector R Martinez-Menchaca[2], BDS, MSc; Mohammad Ahmed[3], MSc, DDS; Nada Farsi[4], BDS, MSc, PhD

[1]Department of Pediatric Dentistry, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia

[2]Department of Comprehensive Dentistry, School of Dentistry, University of Louisville, Louisville, KY, United States

[3]Dubai Health Authority, Dubai, United Arab Emirates

[4]Department of Dental Public Health, Faculty of Dentistry, King Abdulaziz University, Jeddah, Saudi Arabia

Corresponding Author:
Deema Farsi, BDS, MSc
Department of Pediatric Dentistry
Faculty of Dentistry
King Abdulaziz University
PO Box 80200
Jeddah, 21589
Saudi Arabia
Phone: 966 16402000 ext 20388
Email: dfarsi@kau.edu.sa

Related Article:

Companion article: https://www.jmir.org/2021/4/e23205/

Abstract

Background: People are now connected in a borderless web-based world. The modern public, especially the younger generation, relies heavily on the internet as the main source of health-related information. In health care, patients can use social media for more tailored uses such as telemedicine, finding a provider, and for peer support.

Objective: The aim of this narrative review is to discuss how social media has been used in the health care industry from the perspective of patients and describe the main issues surrounding its use in health care.

Methods: Between March and June 2020, a review of the literature was conducted on PubMed, Google Scholar, and Web of Science for English studies that were published since 2007 and discussed the use of social media in health care. In addition to only English publications that discussed the use of social media by patients, publications pertaining to ethical and legal considerations in the use of social media were included. The studies were then categorized as *health information*, *telemedicine*, *finding a health care provider*, *peer support and sharing experiences*, and *influencing positive health behavior*. In addition, two more sections were added to the review: *issues pertaining to social media use in health care* and *ethical considerations*.

Results: Initially, 75 studies were included. As the study proceeded, more studies were included, and a total of 91 studies were reviewed, complemented by 1 textbook chapter and 13 web references. Approximately half of the studies were reviews. The first study was published in 2009, and the last was published in 2021, with more than half of the studies published in the last 5 years. The studies were mostly from the United States (n=40), followed by Europe (n=13), and the least from India (n=1). WhatsApp or WeChat was the most investigated social media platform.

Conclusions: Social media can be used by the public and patients to improve their health and knowledge. However, due diligence must be practiced to assess the credibility of the information obtained and its source. Health care providers, patients, and the public need not forget the risks associated with the use of social media. The limitations and shortcomings of the use of social media by patients should be understood.

(J Med Internet Res 2022;24(1):e30379) doi:10.2196/30379

KEYWORDS

social media; social networking; internet; health care; COVID-19; patient; telemedicine; mobile phone

Introduction

Background

There has been an inexorable increase in digitization over the last 2 decades. Over the years, internet use has remarkably developed, in a way that its use has become effortlessly easy. Websites have been developed into user-friendly apps, mobile phones have become smartphones, and internet coverage has become broader than ever. Interactive websites (Web 2.0) are increasingly overshadowing traditional static websites. Web 2.0 is a term that refers to different types of websites and applications that allow any user to generate content and share it on the web in a web-based community. Social media is a type of Web 2.0 that has been recently introduced as internet-based websites and apps, where user-generated content is created and conveniently exchanged with other users [1]. It is designed as a space for people to obtain information, share experiences, build communities, connect electronically both informally and professionally, and link them to others with common interests, which led to the emergence of the term *self-media*. Users generally need to create a profile or account on the vector and then determine with whom to share it, whether it is a list of known users with similar interests or a broader public community that has access to the vector.

Research in Context

As the consumption of social media has grown, it has become an essential tool used in many industries. In health care, traditional services have been complemented by social media. A simple search on PubMed with the words *social media* would yield several studies, reflecting how relevant the topic is to health care. Although the vast majority of studies investigated social media from the perspective of a health care provider (HCP), there is an abundance of studies that investigated how patients and the public are using it as a resource to supplement traditional health care. Studies varied in their aims, designs, and methodology, and presented mixed findings. Although most studies found promising results, some findings highlighted several limitations and negative issues regarding the use of social media by patients [2-8]. Most included reviews have focused on 1 or 2 main domains of the use of social media in health care such as telemedicine and smoking cessation [9,10]. To our knowledge, no review has holistically discussed the use of social media from the perspective of a patient. In this narrative review, we try to answer the question, “In what ways have patients used social media in relation to health care?” by accumulating, summarizing, and reorganizing findings from published literature.

Objectives

This review aims to discuss how social media has been an essential tool in the health care industry from the perspective of patients. The discussion is supplemented with a discussion on issues pertaining to the use of social media and the ethical considerations that emerged from the literature.

Methods

Methodology Overview

This review is a continuation of the findings presented in *Social Media and Healthcare, Part 1: Literature Review of Social Media Use by Health Care Providers*, which discussed the use of social media in the health care industry from the perspective of an HCP [11]. The original plan was to conduct a general review on the use of social media in health care. Owing to the abundance of information, a decision was made to divide the findings into 2 reviews.

Search Strategy and Information Sources

In the first phase, a comprehensive search on PubMed, Google Scholar, and Web of Science was conducted in March and April 2020 for medical publications on the use of social media in health care in English from 2007 to date. A combination of the following keywords was used to search for relevant articles: *social media* (Medical Subject Headings [MeSH] term) OR *social networking/social network* OR *internet* (MeSH term) OR *Instagram* OR *Facebook* OR *WhatsApp* OR *LinkedIn* OR *YouTube* OR *Twitter* AND *health care* OR *health* (MeSH term) OR *medicine* (MeSH term) OR *physician* (MeSH term) OR *nursing* (subheading) OR *dentistry* (MeSH term) OR *telemedicine* (MeSH term), *recruitment*, OR *education* (subheading) OR *career* OR *behavior/behaviour* (MeSH term) OR *research* (MeSH term). As studies emerged, a second search was conducted in June 2020 with the following combinations: *social media* (MeSH term) OR *social networking* OR *internet* (MeSH term) AND *legal liability* (MeSH term) OR *professionalism* (MeSH term) OR *impact* (MeSH term) OR *ethics* (MeSH term) OR *limitation* OR *harm*.

Screening Process

An EndNote (EndNote 20; Clarivate Analytics) library was created, in which the articles were entered and duplicate publications were removed. For articles to be included, they had to (1) be about social media and health care from the perspective of patients; (2) be in the English language; (3) have accessible full text; and (4) be published in 2007 or later. Exclusion criteria were as follows: (1) abstracts only, without full text; (2) non-English; and (3) irrelevant, such as those discussing social media use from the perspective of an HCP or the use of non-Web 2.0 applications. Reviews and observational and experimental studies were included, with no exclusion based on the study design. The eligibility of the titles and abstracts was also assessed. Finally, the full texts were retrieved. Manual reference screening of the included studies was performed to locate other relevant articles.

Categorization

On the basis of the key outcomes, articles were initially divided into two groups: *patient/the public* and *other relevant issues*. As more information was obtained, the latter was further divided into two groups: *issues pertaining to social media use in health care* and *ethical considerations*. *Issues pertaining to social*

media use in health care covered studies on the limitations, negative effects, and harms of use of social media in health care that emerged from the literature. *Ethical considerations* presented information about legal and ethical issues pertaining to the use of social media in health care.

To best present the findings, the group titled *patient/the public* was subsequently divided into 4 subgroups. The first subgroup was *health information*; although this point was discussed in the first review, in this part we have discussed how patients receive information, rather than how HCPs disseminate it. The second subgroup was *telemedicine*; issues pertaining to the use of telemedicine by patients were discussed. *Finding an HCP* was the mirror image of the group named *career development/practice promotion*, which was discussed in the first review. In the previous review, we discussed how HCPs use social media to market themselves and their practice, whereas in this study, we explored the impact of this on patients' decision-making. The fourth subgroup was *peer support and sharing experiences*, which was unique to patients and the public, and discussed how social media is used among patients for compassion and as a digital word of mouth.

In the first review, a section titled *influencing positive health behavior* was comprehensive. After reviewing it, a decision was made to move it to this review as a fifth group, as it was more relevant to patients than HCPs.

Results

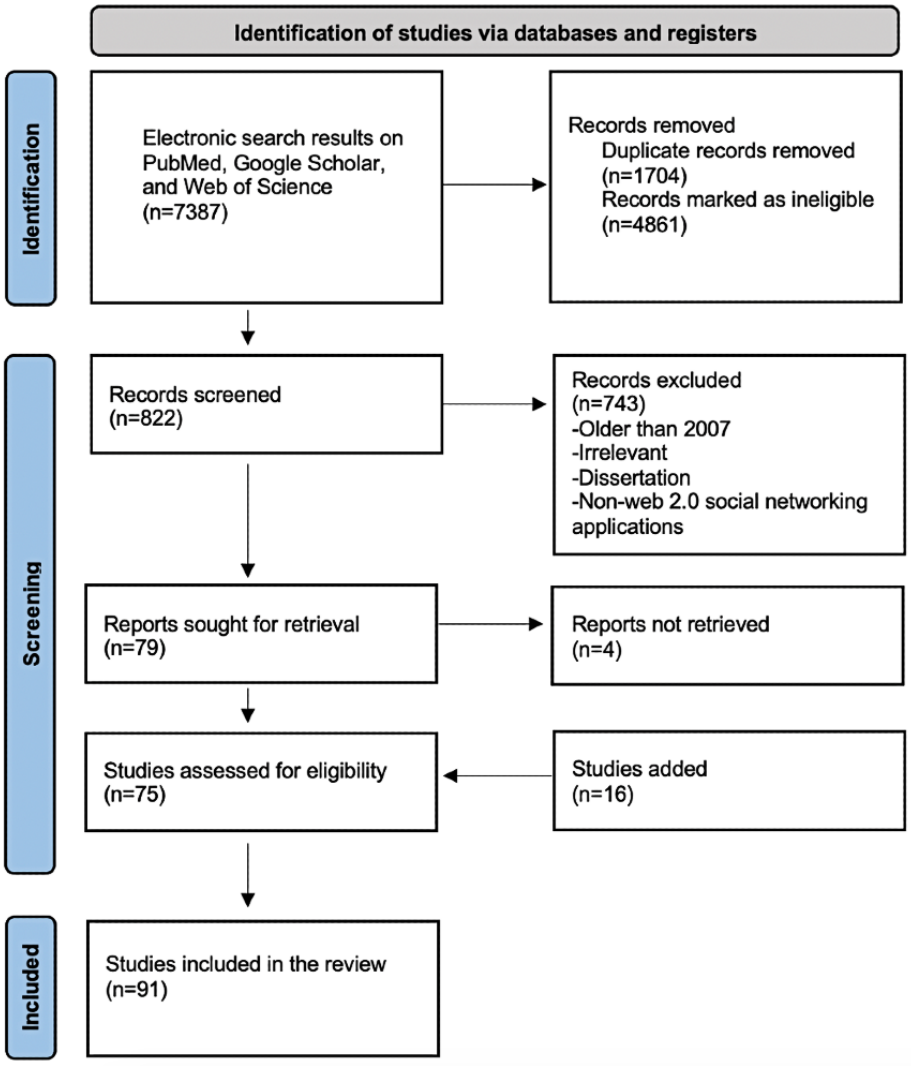
Overview

In this section, the search results in terms of the included publications are presented. The findings pertaining to the content of the individual studies were categorized and are presented in the *Discussion* section.

Search Results

A total of 7387 articles were retrieved from the search, and after removing the duplicate articles, 5683 (76.93%) articles remained. A total of 85.53% (4861/5683) of articles were marked as ineligible and were thus excluded. An additional 13.07% (743/5683) of articles were excluded after title and abstract screening based on the inclusion/exclusion criteria, and 0.07% (4/5683) were irretrievable. The full text of 1.31% (75/5683) of publications was screened and included. Owing to the daily emergence of relevant publications and reference screening, 16 more studies and 1 textbook chapter were added as the review proceeded by updating the search. A total of 91 articles and 1 textbook chapter were included in the analysis. Figure 1 shows a flow diagram explaining how the final inclusion was attained after the selection procedure.

Figure 1. Flowchart of the literature search results.

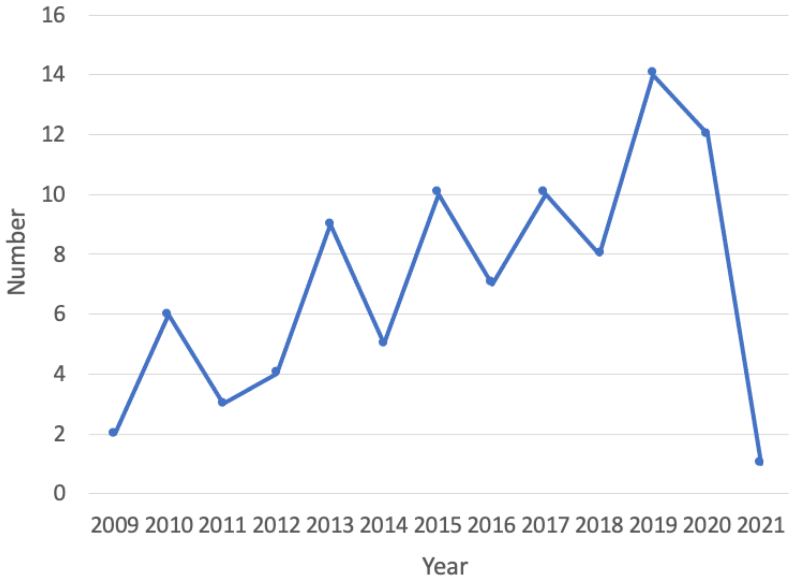


Characteristics of Included Studies

Figure 2 shows the number of included studies per publication year, with more than half of them published in the last 5 years. In terms of geographic location, the 91 publications were distributed as follows: 40 (43%) from the United States, 6 (6%) from Canada, 2 (2%) from Latin America, 10 (10%) from the United Kingdom, 13 (14%) from Europe, 8 (8%) from the Middle East, 1 (1%) from India, 7 (7%) from Asia, and 4 (4%) from Australia.

The included publications were complemented with web references and a textbook chapter. Original studies accounted for 42.8% (39/91) of the cited references. The remaining publications were meta-analyses, systematic reviews, narrative reviews, coping reviews, short communications, commentaries, viewpoint papers, and overviews. The social media platforms specifically investigated in some of the studies were Twitter or Weibo (n=1), WhatsApp or WeChat (n=10), Facebook (n=6), YouTube (n=2), Instagram (n=3), and blogs (n=1). Multimedia Appendix 1 [1-91] provides characteristics of the included 91 studies in chronological order.

Figure 2. Number of included publications per year.



Qualitative Synthesis of the Results

All relevant information regarding the research question was extracted and summarized from the included studies. Information was then categorized into the emerging themes, as presented in the review: (1) social media use from the perspective of patients; (2) issues pertaining to the use of social media in health care; (3) ethical considerations; and (4) public health implications. The retrieved information was then qualitatively synthesized in the discussion for each category.

Discussion

Principal Findings

HCPs and patients typically represent the 2 ends of most health care relationships. HCP is a term used in this review to include physicians, dentists, nurses, medical or dental allied personnel, and health care organizations, whereas patients is a term used to include patients under the care of an HCP and the public. There is overlap in the ways HCPs and patients use social media. In the following section, only information unique to the perspective of patients, which has not been covered in Part I, is presented [11]. Collaterally during the search, studies that investigated ethical and legal considerations in the use of social media and others that discussed its shortcomings and barriers have emerged. These points have also been briefly discussed.

Social Media Use From Patients' Perspective

Overview

In this digital age, people are accustomed to using the internet for health communication. The new term *netizen* has been introduced and is informally used to describe a habitual user of the internet. It is indisputable that patients greatly incorporate social media in seeking health care and that the public is heavily reliant on it to obtain health care information. Perhaps no example supports this notation, as recently witnessed amid the COVID-19 pandemic. There is an abundance of information in the literature pertaining to this subject. In the following section, information has been presented in 5 categories.

Health Information

For a good proportion of the public, young people in particular, social networking sites are the first resource to find general and health-related information [1]. Many individuals with a medical concern are now seeking answers on the web and can virtually obtain them at anytime from anywhere [12]. Social media has radically transformed the way patients obtain information about procedures as well. In a 2009 study, 61% of American adults reported looking on the web for health information [13]. Another study in 2013 found that the first motive of patients for health-related use of social media is seeking information about health, a disease, or treatment of a disease; Twitter was the most commonly used platform for that information [14]. Moreover, 74.9% of web-based health-related information seekers searched for oral health–related information [15].

Health organizations, HCPs, and lay people make an exceedingly large amount of health-related information available on social media. However, the amount of information available may be overwhelming, and the sources may be unverified. The authenticity of the information posted should be questioned, and the recipients must be wary of the information they encounter because many posts do not undergo any quality regulation or verification, and the users are usually in control of the content they encounter [13].

Perhaps there has never been a time where social media was used to obtain health care information, as was the case during the COVID-19 pandemic. In a single day in March 2020, COVID-19–related terms were mentioned more than 20 million times on social media [92]. Almost every social media platform imaginable contributed to the dissemination of information pertaining to the pandemic. Health authorities have used their social media accounts to effectively share scientific information and combat what has been described as an infodemic [93]. Now that vaccines against SARS-CoV-2 are available, social media has been used again as a public podium for individuals to share their thoughts of and experiences with vaccination. Although social media has an unprecedented capacity to make evidence-based information accessible to the public and promote

positive health behaviors, it has also been a major factor in propagating vaccination hesitancy, thus posing a threat to global public health [16,17].

In conclusion, HCPs will continue to be challenged by misinformation readily available to patients on social media. They must be determined to abide by evidence-based health care and ready when challenged by misinformed patients. HCPs also have a duty to make scientifically solid information more accessible to the public. At present, targeted health education interventions are strongly encouraged to foster public trust in vaccination and increase their uptake of the COVID-19 vaccine.

Telemedicine

Communication and monitoring in health care have been outsourced to social media in recent years. Appointments became web-based, health information became available on the internet, and examinations and laboratory results became available on the web-based portal of the facility [18]. Care has been delivered remotely through telemedicine apps, which are the best access to care for some populations, such as those in isolation or in rural areas [12]. Monitoring patients in their homes can improve health care services [19]. Good overall satisfaction has been reported with new telemedicine strategies that shift care to a more patient-centered one [9]. Not only is telemedicine efficient, but it is also time- and cost-saving.

In a 2016 study, telemedicine impression via WhatsApp and clinical assessments was consistent in 82% of the cases examined. Furthermore, telemedicine consultation reduced geographic barriers for initial clinical consultations, and most patients were encouraged to pursue a clinical examination [4]. For instance, Georgia Health Sciences University has enabled patients to access a web-based platform to reach their physicians to ask questions or request prescription refills [20]. There is evidence that telemonitoring of pregnancy is effective, especially for patients in rural areas who do not have to travel to a hospital [9]. In a 2018 study on telemedicine in China, a participant made a comment that suggested seeing a physician while staying at home if people could shop while staying at home [21].

To summarize, patients are encouraged to use telemedicine services that have become readily available and have remarkably improved since the COVID-19 pandemic. However, they must also remember that telemedicine is not the only means to receive health care, nor is it suitable for all cases. Patients have a right to traditional health care as needed and must comply with traditional appointments and hospital visits that are deemed necessary by the treating physician.

Finding an HCP

Social media has now become the new word of mouth. Web-based resources are being increasingly used and highly regarded to make health care decisions, including finding an HCP [6]. In fact, a considerable number of patients are currently searching for HCPs on social media. Some make educated decisions after comprehensive research on the academic qualifications and experience of the practitioner, whereas others follow their emotions after encountering an inviting post or an attractive image, with the latter comprising a huge pool of patients [6,22,23,94].

The content available on social media has an impact on prospective patients: 41% of social media users are influenced by the content they encounter on the web [95]. For example, a study showed that patients are keen to know qualifications of dentists before they visit the office and may use of LinkedIn for that purpose because many dentists showcase their expertise on that platform [24]. Furthermore, patients ranked academic qualifications as the most important content they sought on a Facebook page; some reported that they also sought positive reviews and awards in addition to the original content. In another study, patients reported that the most important factors in selecting a dentist on social media were the reviews and the qualifications of the dentist, with the least important factors being the awards obtained and the number of likes [22].

The attractiveness of a practitioner or provider on social media should not be underestimated. In fact, a study found that 57% of consumers thought that hospitals' social media presence would strongly influence their hospital choice [25]. In another study, 53.4% strongly agreed about the necessity of having a social media presence for dental practices, and 55.1% thought that social media presence was effective in attracting new patients [22]. An interesting study on plastic surgery practices found that the average total number of followers per practice was significantly associated with the placement of the practice on the front page of Google, compared with the second page. Even after a multivariate adjustment of years of experience and education, use of social media remained an independent predictor of placement on the front page of a Google search [6]. A review by Nayak and Linkov [26] showed that patients used social media to find surgeons and that the social media presence of the surgeon can dramatically increase their image as an expert. On the other hand, it was found that unprofessional behavior of an HCP on social media can adversely affect the trust of patients [27,28].

Similar to most marketing strategies, there is no one-size-fits-all means to be successful as an HCP on social media. However, if HCPs recognize the importance of building a relationship with their audiences through social channels, their brands will become more credible and appealing to the target patients. On the other hand, patients must perform due diligence to profile HCP credentials and not rely solely on their perception of their presence on social media.

Peer Support and Sharing Experiences

Not only do HCPs find support and compassion on social media but also do patients. Individuals with chronic disease use social media to communicate with others and exchange experiences. This is especially helpful in rare medical conditions, in which case patients may be geographically distant. Even the family and friends of patients can receive emotional support or request guidance and advice from health care professionals on social media platforms.

Facebook groups for individuals with specific medical conditions are abundant and actively engage members in peer-to-peer support [29,30]. A number of social networking sites, such as *PatientsLikeMe*, provide patients with information and the opportunity to gain support from other people with the same medical condition [31]. Instagram accounts have also been

created to provide information and peer-to-peer support for patients with health care needs, such as adolescents with type 1 diabetes [32]. Moreover, a study showed that a WhatsApp group for hypertensive patients with type 2 diabetes promoted the adherence of patients to treatment [33].

Health-promoting messages coming from social networks instead of experts were perceived as less disempowering and more effective [13,34]. YouTube has been used by patients with cancer to share personal stories [35]. Moreover, a recent study explored cancer survivorship on social media and found that the content shared by survivors displayed their physical, emotional, and psychological health [36]. Although Instagram was used mainly for sharing images posted by survivors themselves or others, Twitter was used primarily for sharing facts and fundraising. In the first week of the COVID-19 pandemic, Twitter users were found to use the tool to notify or warn their friends and followers about the outbreak; that is, Twitter was a platform for people to bond around the topic of COVID-19 [37].

Patient experience is receiving a substantial amount of attention lately, and social media provides patients with opportunities for their voices to be heard and their conversations to be amplified. They can share their experiences in discussion forums, via instant messaging, or post them on the web for the public to see [38]. As patient communities become more interconnected, patients can recommend or defame a practice and compare different experiences. Social media also allows patients to *like* posts, which may elicit notifications to others in their networks [39]. Word-of-mouth marketing between patients with similar conditions or circumstances is also easy with social media. Recommendations or opinions of users have been perceived to be more credible than other advertisement methods, mainly because of the personal nature of the communication that takes place between users on social media [13].

In conclusion, patients find support from peers on social media and express their feelings about their well-being and the health care they receive. It seems that a snowball effect occurs in patient communities on social media, where the more patient-generated content is being shared, the more the public is attracted, the more interaction takes place, and the more content is generated in return.

Influencing Positive Health Behavior

Supplemental electronic communication with patients has been found to emphasize health care guidelines and improve treatment adherence in patients with chronic diseases [40]. In 1 study, 60% of physicians reported favoring interacting with patients on social media to encourage behavioral changes and drug adherence in the hope that these efforts would lead to better health outcomes [41]. Through social media platforms, HCPs can disseminate positive messages to a wide population of users swiftly and influence healthier behaviors through social reinforcement [42]. For example, a study used several social media platforms to encourage blood donation, indicating that social media helped to improve blood donation practices in Saudi Arabia, where there is a shortage of blood donors [2]. Furthermore, a 23-fold increase in donor pledge in web-based state organ-donor registries was observed just a week after Facebook allowed its users to state their organ-donor status in their profile [42]. A review by van den Heuvel et al [9] found that exercise apps possibly led to less gestational weight gain and an increase in smoking abstinence in pregnant women.

Social media can also increase the public's awareness and compassion toward individuals with special health care needs. Social media platforms are increasingly being used for antistigma campaigns to influence public attitudes. Having their unheard voices made public without barriers can be of tremendous relief to individuals with special health care needs. An example is the role of social media in destigmatizing epilepsy [43]. Moreover, Twitter has been successfully used to combat mental illness stereotypes. The platform has facilitated education and contact between individuals with mental illness and has also highlighted injustice [44]. Facebook also enables users to discuss mental illness without the burden of social discomfort [44]. In China, where sharing the intention to attempt suicide on social media is considered a public health concern, social media can be successfully used to enhance suicide literacy and thus be effective for reducing the stigma attached to suicidal ideation and increasing help-seeking behaviors [45]. In Australia, social media is considered an effective means of delivering suicide prevention activities to a large number of young adults [46]. A project called #chatsafe was developed to assist young people in communicating about suicide via social media to feel better and deglorify suicide; the project was recently globalized [47,48].

Just as social media has the potential to promote healthy behaviors, it can also reduce risky behaviors. It can expand the reach of public health efforts and deliver intervention content in an interactive format. An example is smoking cessation campaigns [49]. Reminders and discussions on Facebook and WhatsApp were found to be effective in preventing smoking relapse in individuals who had stopped smoking [50]. In a 2017 systematic review, Facebook and Twitter were found to be feasible and preliminarily effective for smoking cessation, with studies reporting greater abstinence, reduction in relapse, and an increase in quitting attempts among users [10]. These findings are in agreement with the results of a more recent review, in which the use of Facebook, Twitter, and WhatsApp by an online smoking cessation community showed promising results in helping smokers quit [51]. An initiative on Facebook targeted young adults as an intervention for smoking and heavy drinking [52]. Although the interest in changing smoking habits was bigger than that for drinking behavior, and the participants favored changing 1 habit at a time, they accepted and received the post messages well. In a review by Kazemi et al [53], social media was found to help provide HCPs with a platform for combating illicit drug use. It was also found that social media can identify patterns of emerging drug use and that data mining tools can complement the current surveillance methods for tracking drug abuse. In a 2019 cross-sectional study, Generation Z and millennials, a population with high rates of substance use disorder, thought that social media platforms could be helpful in preventing recurrent drug use; however, fewer than half of the participants expressed a willingness to be monitored via social media to support their recovery [54]. Participants from both cohorts had seen more drug cues on social media than they

saw recovery information, which highlights the need for digital interventions to improve drug use treatment and recovery outcomes.

The impact of social media on sexual behavior has also been investigated. One study created an intervention page on Facebook to promote sexual health and serve as a safe space for youth to share ideas and experiences with peers and professionals [55]. It was reported that for a short term (baseline to 2 months), condom use among high-risk youth in the intervention group was stable, whereas it decreased in the control group. Furthermore, the Facebook initiative was able to reach minority communities in which sexually transmitted infections and HIV infections were prevalent. In a 2016 review, 51 studies that investigated social media for sexual health promotion with social media as the sole intervention or in combination with other interventions were reviewed [56]. A total of 8 publications reported increased condom use, use of health services, and HIV self-testing. Two publications reported a reduction in gonorrhea cases and an increase in syphilis testing. Most publications targeted the youth. Facebook is the most commonly used social media platform, either exclusively or in conjunction with other platforms.

There is evidence that social media promotes physical activity and weight loss. In China, a study compared weight loss among participants in a control group (receiving routine publicity on weight loss) and those in a WeChat group with 6 months of weight loss intervention [3]. Male participants in the WeChat group lost significantly more weight than their control peers, although the former were significantly younger. It was found that the more actively participants were using WeChat, the more weight they lost. Another study among medical students found that those who were part of a motivational Facebook group increased their physical activity after 1 month. The likelihood (odds ratio) of becoming sufficiently active by joining the Facebook page was 3.51 [57]. A study on 341 college students with obesity found that the social media approach facilitated short-term weight loss, with the participants losing considerable weight at 6 and 18 months [58]. An initiative on Instagram was found to be attractive and effective in reinforcing the maintenance of an appropriate level of physical activity [59]. In another study, a health app was developed and found to be successful in motivating users to be physically and socially active in real life [60]. During the COVID-19 pandemic, videos of trainers motivating people to work out in their homes during the lockdown went viral. Similar initiatives were seen taking place on every continent, and what could have been a depressive sedentary lockdown to many became a more bearable time.

Cancer prevention efforts have traditionally focused on adults. As health behaviors can aid in cancer prevention, and many behaviors are established in young adulthood, it is logical to target preventive programs in the younger population. In addition, because most of today's youth are digital natives, using social media for promoting cancer-preventing behaviors seems to be a promising strategy. A comprehensive study discussed the potential of social media in cancer prevention and laid the foundation for future research [61].

A comprehensive 2019 systematic review found variation in the strength of evidence regarding the impact of social media on behavior change [96]. However, social media campaigns have generally aided in the reduction of sedentary behavior, contribution to smoking cessation, and improved sexual health, in addition to being cost-effective. It was also found that social media better prompted users to access support services, especially smoking quit phone services. Illicit drug and smoking campaigns appeared to be more effective for the younger generation. Furthermore, expanding the duration or intensifying campaigns was found to be effective. Evidence suggests that targeting messages at a specific target audience increases their impact.

In conclusion, social media has helped patients adhere to treatment, access health care guidelines, and adopt positive health habits to varying degrees. There is no single platform for obtaining these positive outcomes. Stakeholders, researchers, and HCPs must use the platform they consider more effective for and accessible by their target population and customize their content in terms of simplicity, frequency, method, and duration. Researchers should aim to conduct studies that can be effectively adapted to more than one platform or setting and reach a larger population. Future studies should include greater racial diversity among the participants.

Issues Pertaining to Social Media Use in Health Care

There will always be a positive and a negative side of using social media in health care [62]. Although social media has been heavily used by health organizations, medical personnel, patients, and the public, in general, its use is associated with barriers, limitations, and shortcomings. First, internet connectivity is required to access social media. Despite the widespread use of the internet worldwide, 41% of the global population still has no access to the internet [97]. Unfortunately, low-income families and individuals with disabilities are less likely to use the internet, resulting in further exclusion of individuals who are already marginalized [63] Second, some degree of technology skills is essential to enter the digital world. Although basic skills are not very difficult to acquire, digital literacy can be challenging for some populations, such as older adults and individuals with intellectual impairment [64].

Some studies have investigated the shortcomings of technology-mediated remote health care. Inefficiency of web-based medical visits compared with face-to-face engagements has been perceived [65]. A dermatology study found that the quality of the images obtained in group discussions was inconsistent [66]. There is also a fear that patients enjoying the convenience of telemedicine are deterred from visits to the hospital when necessary [14,67]. Moreover, financial limitations should be considered since e-consultations and web-based visits may not be covered by insurance companies [14].

Connections established through social media may dissolve the boundaries between professional and personal lives [68]. A recent study found that patients often extend internet *friend* requests to their physicians on Facebook; however, recommendations often discourage personal web-based communication between practitioners and patients [40]. Personal

boundaries may be violated by inappropriate curiosity, as social media can provide a wealth of information about its users [25,69]. Patients may have unrestricted access to the personal information of HCPs available on the internet, and HCPs also have access to patient information that may not be available in the health care setting. Nevertheless, patient information received from web-based sources may be helpful in certain health care settings; for example, HCPs may observe a lack of adherence to medical recommendations and may alter management accordingly [18].

In social media communication between patients and HCPs, there may be frequent interruptions; the false sense of *having to* be available 24/7; disparity on urgency; compromised verbal communication and body language, especially in texting services; noncompliance with specific terms of a social media platform; lack of proper guidelines for group moderators to manage discussions and controlling content; difficulty in obtaining printed records of communication; and no accurate records of all web-based encounters in the medical records of the patients [27,70,71]. There is also the possibility of identity theft, since any user can create an account, use any name and profile picture, and claim to be someone else. For instance, the logo of the American Society of Colon and Rectal Surgeons was used by a hospital in a different country to request an endorsement [8].

Social media is a double-edged sword for HCPs. As fast as a positive review travels, so does a negative one. Patients unhappy with a service, payment, treatment outcome, or legal actions may start a war against the practitioner or practice. Teaming up with more keyboard warriors or internet trolls can have a disastrous emotional and professional impact on HCPs. In 2016, a well-respected orthopedic surgeon was awarded US $480,000 in damages for defamation after continual vilification by a patient and her kin through a website and social media. The defamatory material included a fake shaming website that greatly resembled the legitimate business website of the surgeon, on which they referred to him as *the butcher*. Similar materials were posted on a couple of social media platforms such as Facebook, YouTube, and Pinterest [98].

HCPs usually support and defend one another. However, some may find social media a good medium to begin a battle against a competing HCP, justifiable or not. Negative professional criticism, displayed publicly on social media, is a violation of the medical codes of ethics; it expresses ill will and aims to tarnish the image of one's professional colleagues. Destructive negative criticism of colleagues on social media damages the medical profession and its reputation. On a positive note, digital shaming is unlawful in many countries and may lead to legal consequences [99].

Although it comes at a relatively low cost, the volume of information on social media may be overwhelming. In addition, the information can be unreliable, difficult to prove as valid, vary in quality and consistency, outdated, not subjected to peer review, invalid, incorrect, not applicable to all situations, not generalizable, opinions and preferences presented as facts, or entirely false [14,38,72]. This is a public health threat, the effect of which is difficult to quantify. It can be difficult for inexperienced HCPs and the public to discern reliable information; thus, there is a risk of absorbing both valid and less credible information. With digital media, social media in particular, misinformation can be easily amplified within echo chambers, which consist of individuals with similar mindsets and beliefs [73]. With artificial intelligence incorporated into technology, algorithm-driven filters selectively display content based on user preferences [73]. For example, a mother who is uncertain about vaccinating her child may join a group of antivaccine mothers to learn more about their concerns. Not only would she be bombarded with antivaccination information, from that point on, antivaccination related information will be targeting her on several social media platforms, fostering antivaccination which may not be at her nor her child's best interest.

It is a fact that public voices disseminating inaccurate health information are usually far better heard and related to than evidence-based knowledge from experts and official health organizations [74]. It was noted that disinformation travels at the same speed that information does, which is why some organizations and authorities have dedicated time and effort to fight myths and disinformation in social media platforms, as seen in the exclusive website section of the World Health Organization dedicated to myth-busting COVID-19 disinformation [72,75]. Another negative consequence of social media is the poorly defined audience; information shared by HCPs may entirely miss the target population. Moreover, with social media, there is a risk of early adoption of unvalidated research and preliminary findings that carry a risk of future medical reversal, which would create more hesitancy in the public and HCPs alike [73]. Another major problem in publishing scientific information on the web is that the user may have hidden conflicts of interest that are not disclosed. It is crucial that every effort be made to critically appraise the information available on social media.

The rapid speed at which information travels may have a very negative impact on the general well-being of the public. For example, disseminating alarming and exaggerated information, misinformation, and manipulated information about COVID-19 may cause fear, anxiety, undue stress, and depression at a societal level, even in individuals without underlying psychiatric illnesses [72]. People may also publicly share their negative feelings, such as anxiety, worry, and conspiracism on social media. Such posts may have a contagious effect. At the beginning of the COVID-19 pandemic in the first few weeks of 2020, a study in China surveyed over 4000 participants. Frequent exposure to social media was associated with high odds of anxiety and depression in the general population as well as among health care workers [5]. Another study found that 53.8% of respondents expressed encountering a moderate or severe psychological impact from the COVID-19 pandemic [76]. Furthermore, a UK study found a positive relationship between the use of social media as a source of information on COVID-19 and conspiracy theory beliefs, especially among younger participants [77].

Being highly influential and used by a large young population, social media may also promote unhealthy habits such as tobacco and alcohol use, violence, unhealthy dietary choices, and

high-risk sex, especially if they are promoted by digital community leaders (ie, influencers) [70,78,79]. Furthermore, enforced advertisement on social media and the subconscious messages of what *looks* good through seductive photographs may have negative unintended consequences for body image and self-esteem in some users and could provide patients with unrealistic expectations for treatment [80]. The public is usually unaware that practitioners showcase successful outcomes selectively and that the pictures may not reflect the true skills and proficiency of a practitioner [71]. This may discourage students and recent graduates who may have not yet obtained the skills of experienced HCPs. Some social media groups are based on misconceptions and can be misleading to the public, such as groups that promote freedom to take off the masking during the COVID-19 pandemic. However, social media platforms have begun taking action to limit discussions of that sort [74].

Posting photographs of procedures and *before-and-after* photographs in a reasonable amount may be beneficial and educational; however, some practitioners make it a goal in itself. If overdone, these posts lose their educational value and become unprofessional advertising and marketing tools [80]. In addition, the pressure to be socially accepted and celebrated, especially through social media, may be difficult to handle. Some individuals, including HCPs, measure their self-worth and seek validation from feedback on social media (eg, number of followers, retweets, and likes). Social media users whose self-confidence is lacking can become more anxious or depressed, which will lead to less self-confidence and erosion of self-worth. It is advisable that HCPs re-evaluate the value of social media if it starts to affect them negatively. It might be advisable to cut back or opt-out all together. Just as it applies to the public, if HCPs are psychologically impacted and struggling, it is better to seek professional help early on.

Although the use of social media among adolescent patients has been shown to be effective in promoting positive health behaviors such as increased physical activity and smoking cessation, the negative impact of social media on the mental health of young people cannot be neglected [50,60]. There is evidence to support less use of social media as a protective factor for mental health in young people [81,82]. In recent years, cyberbullying has emerged as a threat to the mental well-being of young people. A 2015 review found a consistent relationship between cyberbullying and depression among adolescents [83]. In another review, victims of cyberbullying were found to be affected by worry, fear, depression, and loneliness [7]. It was also found that being a cyberbullying victim was associated with more self-injurious behaviors and suicidal thoughts. In the 2019 study by Viner et al [84], the authors analyzed data from the Longitudinal Study of Young People in England and found that the frequent use of social media by young girls was associated with decreased well-being and increased psychological distress. However, they also found that the negative impact of using social media appears to stem from the harmful content users are exposed to and the displacement of healthy lifestyles rather than social media use per se. A review in 2017 found that social media use substituted social interactions, leading to depression and anxiety [7].

A major problem with social media use is that the content posted is prone to be judged and evaluated by whoever sees it. The judgment can be very subjective based on the rater and may reflect unfavorably on HCPs. The trust of patients may be shaken over one *bad* or *inappropriate* post. There are no clear guidelines about e-professionalism and what is considered appropriate; it is inherently subjective [85,86]. A review by Neville and Waylen [27] displays practical examples of e-professionalism that help simplify the concept. The digital footprint has an impact not only on the reputation of the user but also on the profession. Postings on social media can be a permanent record, even after the content is deleted.

Social media posts can be viewed by a large audience base beyond the intention or imagination of users [38]. Employers, program directors, and health officials have the authority to discipline HCPs for unprofessional behavior or breaches of patient privacy, which may ultimately affect the credentials and licensure of the practitioners [20,40,87]. Even appropriate posts may be unfairly scrutinized and negatively judged when viewed out of context. There is also the problem of conflicting timestamps, such as a tweet or a post shared at a time when the HCP was in the middle of a procedure or should have given greater attention to a clinical situation, which could be very damaging to a jury of peers and the public's opinion [8].

In the United Kingdom, 45% of pharmacy students stated that they have posted content on the web about which they are not comfortable with future employers seeing [88]. In addition, about 60% of medical schools reported incidents in which students posted inappropriate content on the web [20]. Furthermore, over half of the medical students surveyed in one study admitted to having embarrassing Facebook photographs of themselves [89]. In a study by Langenfeld et al [86], 12.2% of residents had had clearly unprofessional behavior on Facebook, such as Health Insurance Portability and Accountability Act violations and binge drinking; an additional 14.1% demonstrated potentially unprofessional behavior, including political statements and the use of alcohol and tobacco [86].

Ethical Considerations

Social media communications with or about patients can lead to a breach of privacy and anonymity of patients, which may result in legal actions against HCPs and their institutions. To avoid legal consequences, any post about patients, whether in text, video, or image, should be deidentified, in accordance with Health Insurance Portability and Accountability Act regulations [25]. It is advisable to always obtain consent before sharing any patient information, even if the content is anonymized [71,90]. In 2011, an emergency physician discussed patient care on Facebook. Although she did not identify the patient, she shared enough information to make identification easy to others in her community. As a result, she was fired [100]. In 2016, a pediatric anesthesiologist made inappropriate political comments on Facebook and was ultimately fired from the University of Colorado [101].

It is paramount that HCPs realize that professional demeanor is expected on the internet as in real life. Although no formal contract is established between HCPs and patients in the

web-based world, the same rights and responsibilities traditionally applied should be considered on the internet. In 2013, an obstetrician made unsympathetic comments about an always late patient. She accidently made them public. The post and subsequent comments became viral and was featured on the news. Thousands of people petitioned, and the physician endured several professional and personal consequences, but she was not fired from her practice [102]. In another instance, a patient complained to the media about a hospital in California; in retaliation, the hospital disclosed information about the patient to the media without permission and was ultimately fined US $275,000 [103]. In April 2020, an emergency physician in Washington was fired after criticizing his hospital for its COVID-19 response on social media [104].

There are several other issues pertaining to ethical considerations when using social media in health care. One example is the recruitment of minors on social media for research purposes. It is not difficult to locate and recruit research participants below the age of 18 years on social media. However, individuals below that age have not reached cognitive maturity to make thought-through decisions regarding participation in research. Obtaining parental consent or targeting parents may be a more ethical alternative [18]. Another example is falsifying images posted on social media. Photographic technique artifices, such as modifying angles or digitally altering photographs to exaggerate treatment outcomes, is deceiving to patients and is considered unethical abuse [80].

As the use of social media by HCPs has increased, health authorities have published guidelines and recommendations for the use of social media. For example, in 2011, the American Medical Association published its policy on professionalism on social media [91]. Later in 2013, the General Dental Council in the United Kingdom published a document titled, *Guidance on using social media* [105]. It is imperative that medical curricula tackle e-professionalism, professional internet etiquette, and digital ethics, as the use of social media in health care is the new norm among the millennial generation of HCPs. For more information, it is recommended to read the review by Langenfeld and Batra [8], in which recommendations for e-professionalism have been proposed. In addition, refer to the guidelines on the use of social media that have been summarized by Dhar [71].

Public Health Implications

Social media has the potential to transmit health-related information and promote health to the public. Striking the right balance between digital and traditional health care is imperative. Social media is omnipresent in our lives today, and the best guard we have is to be acquainted with it and practice due diligence in using it to our favor for the promotion of health care. Nevertheless, HCPs, patients, and the public in general need not forget the risks to which they may be exposing themselves. As medical professionals, HCPs are bound to ethical principles toward their colleagues, patients, and the public in the digital as much as in the real world. Whether e-professionalism is formally taught, ethics is a matter of choice.

Limitations

Despite its comprehensiveness, because of this review being a narrative review, it is descriptive in nature and did not include a formal appraisal of the included studies. Data from the included studies were summarized and reorganized but not analyzed. Although our search was comprehensive, some relevant studies may have been unidentified. Bias may have occurred in selecting and assessing the literature, as it was not done in a systematic manner, giving the type of review.

Conclusions

This narrative review aimed to discuss how patients have been using social media in the context of health care and describe the main issues pertaining to its use in health care. As can be seen, multidimensional health care, such as when pairing health care with social media and other forms of communication, has been shown to be very successful. The outcome is maximized when the audience is reached numerous times, in multiple settings, and from various sources. The number of digital natives is increasing and will continue to grow in health care settings. Thus, it is advisable to acknowledge that social media will remain an essential part of health care for many years.

Despite emerging evidence that the use of social media has facilitated health care, it has not and will probably not entirely replace traditional health care. The use of social media is associated with barriers, limitations, and shortcomings that continue to emerge in the literature. To maximize the benefits while minimizing compromise to the care provided and avoiding liability, HCPs and patients must perform due diligence before considering social media in health care and should make educated judgments on a case-by-case basis.

As social media is a relatively recent occurrence, more research is needed to determine its long-term effectiveness and to find the best strategies that would maximize its advantages while limiting its risks. e-Professionalism and the ethical considerations in using social media in health care can be further explored.

Acknowledgments

The authors would like to acknowledge the assistance of Dr Maha Qari during the initial search.

Authors' Contributions

The production of this work is collaborative with some overlapping roles. DF conceived the idea, conducted the search, composed the first draft of the manuscript, and submitted the final work for publication (publication correspondence). HRMM conducted the search and revised and approved the manuscript. MA conducted the search and revised and approved the manuscript. NF designed the study, conducted the search, and revised and approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Characteristics of the included 91 studies in a chronological order.
[DOCX File , 47 KB - jmir_v24i1e30379_app1.docx]

References

1. Jain A, Bickham D. Adolescent health literacy and the internet: challenges and opportunities. Curr Opin Pediatr 2014 Aug;26(4):435-439. [doi: 10.1097/MOP.0000000000000119] [Medline: 24886952]
2. Alanzi T, Alsaeed B. Use of social media in the blood donation process in Saudi Arabia. J Blood Med 2019 Dec;Volume 10:417-423. [doi: 10.2147/jbm.s217950]
3. He C, Wu S, Zhao Y, Li Z, Zhang Y, Le J, et al. Social media-promoted weight loss among an occupational population: cohort study using a WeChat mobile phone app-based campaign. J Med Internet Res 2017 Oct 23;19(10):e357 [FREE Full text] [doi: 10.2196/jmir.7861] [Medline: 29061555]
4. Petruzzi M, De Benedittis M. WhatsApp: a telemedicine platform for facilitating remote oral medicine consultation and improving clinical examinations. Oral Surg Oral Med Oral Pathol Oral Radiol 2016 Mar;121(3):248-254. [doi: 10.1016/j.oooo.2015.11.005] [Medline: 26868466]
5. Gao J, Zheng P, Jia Y, Chen H, Mao Y, Chen S, et al. Mental health problems and social media exposure during COVID-19 outbreak. PLoS One 2020 Apr 16;15(4):e0231924 [FREE Full text] [doi: 10.1371/journal.pone.0231924] [Medline: 32298385]
6. Dorfman R, Mahmood E, Ren A, Turin S, Vaca E, Fine N, et al. Google ranking of plastic surgeons values social media presence over academic pedigree and experience. Aesthet Surg J 2019 Mar 14;39(4):447-451. [doi: 10.1093/asj/sjy285] [Medline: 30346492]
7. Hoge E, Bickham D, Cantor J. Digital media, anxiety, and depression in children. Pediatrics 2017;140(Suppl 2):76-80. [doi: 10.1542/peds.2016-1758g]
8. Langenfeld S, Batra R. How can social media get us in trouble? Clin Colon Rectal Surg 2017 Sep 12;30(4):264-269 [FREE Full text] [doi: 10.1055/s-0037-1604255] [Medline: 28924400]
9. van den Heuvel JF, Groenhof TK, Veerbeek JH, van Solinge WW, Lely AT, Franx A, et al. eHealth as the next-generation perinatal care: an overview of the literature. J Med Internet Res 2018 Jun 05;20(6):e202 [FREE Full text] [doi: 10.2196/jmir.9262] [Medline: 29871855]
10. Naslund JA, Kim SJ, Aschbrenner KA, McCulloch LJ, Brunette MF, Dallery J, et al. Systematic review of social media interventions for smoking cessation. Addict Behav 2017 Oct;73:81-93 [FREE Full text] [doi: 10.1016/j.addbeh.2017.05.002] [Medline: 28499259]
11. Farsi D. Social media and health care, part I: literature review of social media use by health care providers. J Med Internet Res 2021 Apr 05;23(4):e23205 [FREE Full text] [doi: 10.2196/23205] [Medline: 33664014]
12. Chaet D, Clearfield R, Sabin JE, Skimming K, Council on Ethical and Judicial Affairs American Medical Association. Ethical practice in telehealth and telemedicine. J Gen Intern Med 2017 Oct 26;32(10):1136-1140 [FREE Full text] [doi: 10.1007/s11606-017-4082-2] [Medline: 28653233]
13. Levac JJ, O'Sullivan T. Social media and its use in health promotion. Interdiscipl J Health Sci 2010 Feb 10;1(1):47. [doi: 10.18192/riss-ijhs.v1i1.1534]
14. Antheunis ML, Tates K, Nieboer TE. Patients' and health professionals' use of social media in health care: motives, barriers and expectations. Patient Educ Couns 2013 Sep;92(3):426-431. [doi: 10.1016/j.pec.2013.06.020] [Medline: 23899831]
15. Almaiman S, Bahkali S, Alabdulatif N, Bahkaly A, Al-Surimi K, Househ M. Promoting oral health using social media platforms: seeking Arabic online oral health related information (OHRI). Stud Health Technol Inform 2016;226:283-286. [Medline: 27350526]
16. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct 23;5(10):e004206 [FREE Full text] [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]
17. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
18. Denecke K, Bamidis P, Bond C, Gabarron E, Househ M, Lau AY, et al. Ethical issues of social media usage in healthcare. Yearb Med Inform 2018 Mar 10;24(01):137-147. [doi: 10.15265/iy-2015-001]
19. Klaassen B, van Beijnum B, Hermens H. Usability in telemedicine systems-A literature survey. Int J Med Inform 2016 Sep;93:57-69. [doi: 10.1016/j.ijmedinf.2016.06.004] [Medline: 27435948]
20. Chauhan B, George R, Coffin J. Social media and you: what every physician needs to know. J Med Pract Manage 2012;28(3):206-209. [Medline: 23373164]

21. Leung R, Guo H, Pan X. Social media users' perception of telemedicine and mHealth in China: exploratory study. JMIR Mhealth Uhealth 2018 Sep 25;6(9):e181 [FREE Full text] [doi: 10.2196/mhealth.7623] [Medline: 30274969]
22. Alalawi A, Aljuaid H, Natto ZS. The effect of social media on the choice of dental patients: a cross-sectional study in The City Of Jeddah, Saudi Arabia. Patient Prefer Adher 2019 Oct;Volume 13:1685-1692. [doi: 10.2147/ppa.s213704]
23. Nayyar A, Jadi J, Garimella R, Elkins-Williams S, Gallagher K, Kalliainen L, et al. Are you on the right platform? A conjoint analysis of social media preferences in aesthetic surgery patients. Aesthet Surg J 2019 Aug 22;39(9):1019-1032. [doi: 10.1093/asj/sjy238] [Medline: 30239573]
24. Parmar N, Dong L, Eisingerich AB. Connecting with your dentist on Facebook: patients' and dentists' attitudes towards social media usage in dentistry. J Med Internet Res 2018 Jun 29;20(6):e10109 [FREE Full text] [doi: 10.2196/10109] [Medline: 29959108]
25. Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
26. Nayak LM, Linkov G. Social media marketing in facial plastic surgery: what has worked? Facial Plast Surg Clin North Am 2019 Aug;27(3):373-377. [doi: 10.1016/j.fsc.2019.04.002] [Medline: 31280851]
27. Neville P, Waylen A. Social media and dentistry: some reflections on e-professionalism. Br Dent J 2015 Apr 24;218(8):475-478. [doi: 10.1038/sj.bdj.2015.294] [Medline: 25908363]
28. Cain J, Romanelli F. E-professionalism: a new paradigm for a digital age. Curr Pharm Teach Learn 2009 Dec;1(2):66-70. [doi: 10.1016/j.cptl.2009.10.001]
29. Grajales FJ, Sheps S, Ho K, Novak-Lauscher H, Eysenbach G. Social media: a review and tutorial of applications in medicine and health care. J Med Internet Res 2014 Feb 11;16(2):e13 [FREE Full text] [doi: 10.2196/jmir.2912] [Medline: 24518354]
30. Childs L, Martin C. Social media profiles: striking the right balance. Am J Health Syst Pharm 2012 Dec 01;69(23):2044-2050. [doi: 10.2146/ajhp120115] [Medline: 23172261]
31. MacMillan C. Social media revolution and blurring of professional boundaries. Imprint 2013;60(3):44-46. [Medline: 23821819]
32. Yi-Frazier JP, Cochrane K, Mitrovich C, Pascual M, Buscaino E, Eaton L, et al. Using Instagram as a modified application of photovoice for storytelling and sharing in adolescents with type 1 diabetes. Qual Health Res 2015 Oct 22;25(10):1372-1382 [FREE Full text] [doi: 10.1177/1049732315583282] [Medline: 25904674]
33. Ramirez JS. Social networks as a means of monitoring patients with hypertension and diabetes success story / Las redes sociales como medio de monitoreo en pacientes con hipertension y diabetes una historia de exito. Int J Integr Care 2015 Nov 17;15(8):74. [doi: 10.5334/ijic.2321]
34. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]
35. Chou WS, Hunt Y, Folkers A, Augustson E. Cancer survivorship in the age of YouTube and social media: a narrative analysis. J Med Internet Res 2011 Jan 17;13(1):e7 [FREE Full text] [doi: 10.2196/jmir.1569] [Medline: 21247864]
36. Cherian R, Le G, Whall J, Gomez S, Sarkar U. Content shared on social media for national cancer survivors day 2018. PLoS One 2020 Jan 15;15(1):e0226194 [FREE Full text] [doi: 10.1371/journal.pone.0226194] [Medline: 31940384]
37. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top concerns of tweeters during the COVID-19 pandemic: infoveillance study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
38. Moorhead SA, Hazlett DE, Harrison L, Carroll JK, Irwin A, Hoving C. A new dimension of health care: systematic review of the uses, benefits, and limitations of social media for health communication. J Med Internet Res 2013 Apr 23;15(4):e85 [FREE Full text] [doi: 10.2196/jmir.1933] [Medline: 23615206]
39. Eisingerich AB, Chun HH, Liu Y, Jia H, Bell SJ. Why recommend a brand face-to-face but not on Facebook? How word-of-mouth on online social sites differs from traditional word-of-mouth. J Consum Psychol 2015 Jan;25(1):120-128. [doi: 10.1016/j.jcps.2014.05.004]
40. Farnan JM, Sulmasy LS, Worster BK, Chaudhry HJ, Rhyne JA, Arora VM, American College of Physicians Ethics, Professionalism and Human Rights Committee, American College of Physicians Council of Associates, Federation of State Medical Boards Special Committee on Ethics and Professionalism. Online medical professionalism: patient and public relationships: policy statement from the American College of Physicians and the Federation of State Medical Boards. Ann Intern Med 2013 Apr 16;158(8):620-627 [FREE Full text] [doi: 10.7326/0003-4819-158-8-201304160-00100] [Medline: 23579867]
41. Househ M. The use of social media in healthcare: organizational, clinical, and patient perspectives. Stud Health Technol Inform 2013;183:244-248. [Medline: 23388291]
42. George D, Rovniak L, Kraschnewski J. Dangers and opportunities for social media in medicine. Clin Obstet Gynecol 2013;56(3):453-462. [doi: 10.1097/grf.0b013e318297dc38]
43. Lo AS, Esser MJ, Gordon KE. YouTube: a gauge of public perception and awareness surrounding epilepsy. Epilepsy Behav 2010 Apr;17(4):541-545. [doi: 10.1016/j.yebeh.2010.02.004] [Medline: 20236867]
44. Betton V, Borschmann R, Docherty M, Coleman S, Brown M, Henderson C. The role of social media in reducing stigma and discrimination. Br J Psychiatry 2015 Jun 02;206(6):443-444. [doi: 10.1192/bjp.bp.114.152835] [Medline: 26034176]

45. Li A, Huang X, Jiao D, O'Dea B, Zhu T, Christensen H. An analysis of stigma and suicide literacy in responses to suicides broadcast on social media. Asia Pac Psychiatry 2018 Mar 31;10(1):e12314. [doi: 10.1111/appy.12314] [Medline: 29383880]
46. Robinson J, Rodrigues M, Fisher S, Bailey E, Herrman H. Social media and suicide prevention: findings from a stakeholder survey. Shanghai Arch Psychiatry 2015 Feb 25;27(1):27-35 [FREE Full text] [doi: 10.11919/j.issn.1002-0829.214133] [Medline: 25852253]
47. Robinson J, Hill NT, Thorn P, Battersby R, Teh Z, Reavley NJ, et al. The #chatsafe project. Developing guidelines to help young people communicate safely about suicide on social media: a Delphi study. PLoS One 2018 Nov 15;13(11):e0206584 [FREE Full text] [doi: 10.1371/journal.pone.0206584] [Medline: 30439958]
48. Robinson J, Teh Z, Lamblin M, Hill NT, La Sala L, Thorn P. Globalization of the #chatsafe guidelines: using social media for youth suicide prevention. Early Interv Psychiatry 2021 Oct 15;15(5):1409-1413. [doi: 10.1111/eip.13044] [Medline: 32935440]
49. Clauson K, Ekins J, Goncz C. Use of blogs by pharmacists. Am J Health Syst Pharm 2010 Dec 01;67(23):2043-2048. [doi: 10.2146/ajhp100065] [Medline: 21098377]
50. Cheung YT, Chan CH, Lai CJ, Chan WF, Wang MP, Li HC, et al. Using Whatsapp and Facebook online social groups for smoking relapse prevention for recent quitters: a pilot pragmatic cluster randomized controlled trial. J Med Internet Res 2015 Oct 22;17(10):e238 [FREE Full text] [doi: 10.2196/jmir.4829] [Medline: 26494159]
51. Thota R, Divatia J. WhatsApp: what an app!. Indian J Crit Care Med 2015 Jun;19(6):363-365 [FREE Full text] [doi: 10.4103/0972-5229.158288] [Medline: 26195867]
52. Ramo DE, Meacham MC, Kaur M, Corpuz ES, Prochaska JJ, Satre DD. Development of a social media-based intervention targeting tobacco use and heavy episodic drinking in young adults. Addict Sci Clin Pract 2019 Apr 01;14(1):14 [FREE Full text] [doi: 10.1186/s13722-019-0141-9] [Medline: 30940206]
53. Kazemi D, Borsari B, Levine M, Dooley B. Systematic review of surveillance by social media platforms for illicit drug use. J Public Health (Oxf) 2017 Dec 01;39(4):763-776 [FREE Full text] [doi: 10.1093/pubmed/fdx020] [Medline: 28334848]
54. Curtis BL, Ashford RD, Magnuson KI, Ryan-Pettes SR. Comparison of smartphone ownership, social media use, and willingness to use digital interventions between generation Z and millennials in the treatment of substance use: cross-sectional questionnaire study. J Med Internet Res 2019 Apr 17;21(4):e13050 [FREE Full text] [doi: 10.2196/13050] [Medline: 30994464]
55. Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]
56. Gabarron E, Bradway M, Fernandez-Luque L, Chomutare T, Hansen AH, Wynn R, et al. Social media for health promotion in diabetes: study protocol for a participatory public health intervention design. BMC Health Serv Res 2018 Jun 5;18(1). [doi: 10.1186/s12913-018-3178-7]
57. Todorovic J, Terzic-Supic Z, Djikanovic B, Nesic D, Piperac P, Stamenkovic Z. Can social media intervention improve physical activity of medical students? Public Health 2019 Sep;174:69-73. [doi: 10.1016/j.puhe.2019.05.030] [Medline: 31323599]
58. Godino JG, Merchant G, Norman GJ, Donohue MC, Marshall SJ, Fowler JH, et al. Using social and mobile tools for weight loss in overweight and obese young adults (Project SMART): a 2 year, parallel-group, randomised, controlled trial. Lancet Diab Endocrinol 2016 Sep;4(9):747-755. [doi: 10.1016/s2213-8587(16)30105-x]
59. Al-Eisa E, Al-Rushud A, Alghadir A, Anwer S, Al-Harbi B, Al-Sughaier N, et al. Effect of motivation by "Instagram" on adherence to physical activity among female college students. Biomed Res Int 2016;2016:1546013 [FREE Full text] [doi: 10.1155/2016/1546013] [Medline: 27034927]
60. Tufte T, Babic A. Effect of social media in a mHealth application. Stud Health Technol Inform 2017;238:120-123. [Medline: 28679902]
61. Sarkar U, Le GM, Lyles CR, Ramo D, Linos E, Bibbins-Domingo K. Using social media to target cancer prevention in young adults: viewpoint. J Med Internet Res 2018 Jun 05;20(6):e203 [FREE Full text] [doi: 10.2196/jmir.8882] [Medline: 29871850]
62. Barreto JE, Whitehair CL. Social media and web presence for patients and professionals: evolving trends and implications for practice. PM R 2017 May 18;9(5S):98-105. [doi: 10.1016/j.pmrj.2017.02.012] [Medline: 28527508]
63. Laz TH, Berenson AB. Racial and ethnic disparities in internet use for seeking health information among young women. J Health Commun 2013 Feb;18(2):250-260 [FREE Full text] [doi: 10.1080/10810730.2012.707292] [Medline: 23130608]
64. Smith B, Magnani JW. New technologies, new disparities: the intersection of electronic health and digital health literacy. Int J Cardiol 2019 Oct 01;292:280-282 [FREE Full text] [doi: 10.1016/j.ijcard.2019.05.066] [Medline: 31171391]
65. Muntaner-Mas A, Vidal-Conti J, Borràs PA, Ortega F, Palou P. Effects of a Whatsapp-delivered physical activity intervention to enhance health-related physical fitness components and cardiovascular disease risk factors in older adults. J Sports Med Phys Fitness 2017 Feb;57(1-2):90-102. [doi: 10.23736/s0022-4707.16.05918-1]
66. Kaliyadan F, Ashique K, Jagadeesan S, Krishna B. What's up dermatology? A pilot survey of the use of WhatsApp in dermatology practice and case discussion among members of WhatsApp dermatology groups. Indian J Dermatol Venereol Leprol 2016;82(1):67-69 [FREE Full text] [doi: 10.4103/0378-6323.171638] [Medline: 26728815]

67. Kim S. Content analysis of cancer blog posts. J Med Libr Assoc 2009 Oct;97(4):260-266 [FREE Full text] [doi: 10.3163/1536-5050.97.4.009] [Medline: 19851489]
68. Golder S, Ahmed S, Norman G, Booth A. Attitudes toward the ethics of research using social media: a systematic review. J Med Internet Res 2017 Jun 06;19(6):e195 [FREE Full text] [doi: 10.2196/jmir.7082] [Medline: 28588006]
69. Bosslet GT, Torke AM, Hickman SE, Terry CL, Helft PR. The patient-doctor relationship and online social networks: results of a national survey. J Gen Intern Med 2011 Oct 25;26(10):1168-1174 [FREE Full text] [doi: 10.1007/s11606-011-1761-2] [Medline: 21706268]
70. Boulos MK, Giustini D, Wheeler S. Instagram and Whatsapp in health and healthcare: an overview. Fut Internet 2016 Jul 26;8(3):37. [doi: 10.3390/fi8030037]
71. Dhar V. Social media and its implications in pediatric dentistry. Pediatr Dent 2018 May 15;40(3):174-176. [Medline: 29793562]
72. González-Padilla DA, Tortolero-Blanco L. Social media influence in the COVID-19 Pandemic. Int Braz J Urol 2020 Jul;46(suppl 1):120-124. [doi: 10.1590/s1677-5538.ibju.2020.s121]
73. Chan AK, Nickson CP, Rudolph JW, Lee A, Joynt GM. Social media for rapid knowledge dissemination: early experience from the COVID-19 pandemic. Anaesthesia 2020 Dec 31;75(12):1579-1582 [FREE Full text] [doi: 10.1111/anae.15057] [Medline: 32227594]
74. Stellefson M, Paige SR, Chaney BH, Chaney JD. Evolving role of social media in health promotion: updated responsibilities for health education specialists. Int J Environ Res Public Health 2020 Feb 12;17(4):1153. [doi: 10.3390/ijerph17041153] [Medline: 32059561]
75. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):7516 [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
76. Wang C, Pan R, Wan X, Tan Y, Xu L, Ho CS, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 Coronavirus Disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health 2020 Mar 06;17(5):1729 [FREE Full text] [doi: 10.3390/ijerph17051729] [Medline: 32155789]
77. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol Med 2020 Jun 09;51(10):1763-1769. [doi: 10.1017/s003329172000224x]
78. Coates A, Hardman C, Halford J, Christiansen P, Boyland E. Social media influencer marketing and children's food intake: a randomized trial. Pediatrics 2019 Apr;143(4). [doi: 10.1542/peds.2018-2554] [Medline: 30833297]
79. De Benedictis A, Lettieri E, Masella C, Gastaldi L, Macchini G, Santu C, et al. WhatsApp in hospital? An empirical investigation of individual and organizational determinants to use. PLoS One 2019 Jan 11;14(1):e0209873 [FREE Full text] [doi: 10.1371/journal.pone.0209873] [Medline: 30633754]
80. Simplício AH. Social media and Dentistry: ethical and legal aspects. Dental Press J Orthod 2019 Dec;24(6):80-89 [FREE Full text] [doi: 10.1590/2177-6709.24.6.080-089.sar] [Medline: 31994651]
81. Cairns KE, Yap MB, Pilkington PD, Jorm AF. Risk and protective factors for depression that adolescents can modify: a systematic review and meta-analysis of longitudinal studies. J Affect Disord 2014 Dec;169:61-75. [doi: 10.1016/j.jad.2014.08.006] [Medline: 25154536]
82. DeSmet A. Social media and lifestyles in youth mental health promotion. Lancet Child Adol Health 2019 Oct;3(10):664-665. [doi: 10.1016/s2352-4642(19)30233-0]
83. Hamm MP, Newton AS, Chisholm A, Shulhan J, Milne A, Sundar P, et al. Prevalence and effect of cyberbullying on children and young people: a scoping review of social media studies. JAMA Pediatr 2015 Aug 01;169(8):770-777. [doi: 10.1001/jamapediatrics.2015.0944] [Medline: 26098362]
84. Viner RM, Gireesh A, Stiglic N, Hudson LD, Goddings A, Ward JL, et al. Roles of cyberbullying, sleep, and physical activity in mediating the effects of social media use on mental health and wellbeing among young people in England: a secondary analysis of longitudinal data. Lancet Child Adol Health 2019 Oct;3(10):685-696. [doi: 10.1016/s2352-4642(19)30186-5]
85. Chretien KC, Goldman EF, Beckman L, Kind T. It s your own risk: medical students perspectives on online professionalism. Acad Med 2010;85:68-71. [doi: 10.1097/acm.0b013e3181ed4778]
86. Langenfeld SJ, Cook G, Sudbeck C, Luers T, Schenarts PJ. An assessment of unprofessional behavior among surgical residents on Facebook: a warning of the dangers of social media. J Surg Educ 2014 Nov;71(6):28-32. [doi: 10.1016/j.jsurg.2014.05.013] [Medline: 24981657]
87. Lambert KM, Barry P, Stokes G. Risk management and legal issues with the use of social media in the healthcare setting. J Healthc Risk Manag 2012 Apr 23;31(4):41-47. [doi: 10.1002/jhrm.20103] [Medline: 22528403]
88. Hall M, Hanna L, Huey G. Use and views on social networking sites of pharmacy students in the United kingdom. Am J Pharm Educ 2013 Feb 12;77(1):9. [doi: 10.5688/ajpe7719] [Medline: 23459621]
89. Garner J, O'Sullivan H. Facebook and the professional behaviours of undergraduate medical students. Clin Teach 2010;7(2):112-115. [doi: 10.1111/j.1743-498x.2010.00356.x]

90. Fogelson N, Rubin Z, Ault K. Beyond likes and tweets: an in-depth look at the physician social media landscape. Clin Obstet Gynecol 2013;56(3):495-508. [doi: 10.1097/grf.0b013e31829e7638]
91. Shore R, Halsey J, Shah K, Crigger B, Douglas SP, AMA Council on EthicalJudicial Affairs (CEJA). Report of the AMA Council on ethical and judicial affairs: professionalism in the use of social media. J Clin Ethics 2011;22(2):165-172. [Medline: 21837888]
92. Molla R. How coronavirus took over social media. Vox Media. 2020. URL: https://www.vox.com/recode/2020/3/12/21175570/coronavirus-covid-19-social-media-twitter-facebook-google [accessed 2020-10-22]
93. Working together to tackle the “infodemic”. World Health Organization. 2020. URL: https://www.euro.who.int/en/health-topics/Health-systems/e-health/news/news/2020/6/working-together-to-tackle-the-infodemic [accessed 2020-07-23]
94. Infographic: rising use of social and mobile in healthcare. The Spark Report. 2012. URL: https://thesparkreport.com/infographic-social-mobile-healthcare/ [accessed 2020-03-22]
95. Healthcare infographic. Demi and Cooper. 2012. URL: https://demicooper.com/?project=social-mobile-use-in-healthcare [accessed 2020-07-24]
96. Stead M, Angus K, Langley T, Katikireddi S, Hinds K, Hilton S, et al. Mass Media to Communicate Public Health Messages in Six Health Topic Areas: A Systematic Review and Other Reviews of the Evidence. Southampton (UK): NIHR Journals Library; 2019.
97. Worldwide digital population as of July 2020. Statista. 2020. URL: https://www.statista.com/statistics/617136/digital-population-worldwide/ [accessed 2020-07-23]
98. Surgeon awarded $480,000 for online defamation by former patient. Avant Mutual. 2017. URL: https://www.avant.org.au/news/surgeon-awarded-for-online-defamation/ [accessed 2020-07-24]
99. Cyberbullying laws. FindLaw. 2019. URL: https://criminal.findlaw.com/criminal-charges/cyber-bullying.html [accessed 2020-07-24]
100. Conaboy C. For doctors, social media a tricky case. Boston. 2011. URL: http://archive.boston.com/lifestyle/health/articles/2011/04/20/for_doctors_social_media_a_tricky_case/ [accessed 2020-07-23]
101. Denver institutions cut ties with doctor who called Michelle Obama a 'monkey face' on Facebook. Fanverse. 2016. URL: https://tinyurl.com/y89e5e47 [accessed 2020-07-24]
102. OB-GYN, in hot water after posting comment about patient on Facebook. HuffPost. 2017. URL: https://www.huffpost.com/entry/amy-dunbar-obgyn_n_2630823 [accessed 2020-07-24]
103. Doctors fire back at bad Yelp reviews and reveal patients' information online. The Washington Post. 2016. URL: https://tinyurl.com/2p89ckwb [accessed 2020-07-24]
104. An ER doctor lost his job after criticizing his hospital on COVID-19. Now he's suing. NPR. 2020. URL: https://tinyurl.com/yje38h5e [accessed 2020-07-24]
105. Guidance on using social media. General Dental Council. 2016. URL: https://www.gdc-uk.org/docs/default-source/guidance-documents/guidance-on-using-social-media.pdf?sfvrsn=de158345_2 [accessed 2020-07-24]

Abbreviations

HCP: health care provider
MeSH: Medical Subject Headings

Edited by A Mavragani; submitted 14.05.21; peer-reviewed by E Said-Hung, M Mbwogge; comments to author 14.07.21; revised version received 25.08.21; accepted 30.11.21; published 07.01.22.

Please cite as:
Farsi D, Martinez-Menchaca HR, Ahmed M, Farsi N
Social Media and Health Care (Part II): Narrative Review of Social Media Use by Patients
J Med Internet Res 2022;24(1):e30379
URL: https://www.jmir.org/2022/1/e30379
doi:10.2196/30379
PMID:34994706

Review

Digital Intervention Strategies for Increasing Physical Activity Among Preschoolers: Systematic Review

Taren Swindle[1], BA, MSc, PhD; Anwesh B Poosala[2], MD; Nan Zeng[3], PhD; Elisabet Børsheim[1,4,5], PhD; Aline Andres[1,4,5], PhD; Laura L Bellows[6], RDN, MPH, PhD

[1]Department of Pediatrics, University of Arkansas Medical Sciences, Little Rock, AR, United States

[2]Griffin Hospital, Derby, CT, United States

[3]Department of Pediatrics, University of New Mexico, Albuquerque, NM, United States

[4]Arkansas Children's Nutrition Center, Little Rock, AR, United States

[5]Arkansas Children's Research Institute, Little Rock, AR, United States

[6]Division of Nutritional Sciences, Cornell University, Ithaca, NY, United States

Corresponding Author:
Taren Swindle, BA, MSc, PhD
Department of Pediatrics
University of Arkansas Medical Sciences
Little Rock, AR
United States
Phone: 1 8702360997
Email: tswindle@uams.edu

Abstract

Background: Digital interventions are increasingly used to improve health behaviors. Improved access and lower costs (relative to in-person interventions) make such interventions appealing. Specifically, digital platforms may be a promising approach for increasing physical activity (PA) in young children.

Objective: The goal of this systematic review was three-pronged: (1) to determine the quality of studies using digital PA intervention strategies with preschool-aged children (ie, 3 to 5 years old); (2) to assess the efficacy of digital interventions and approaches designed to improve PA in preschool-aged children; and (3) to examine theoretical application and implementation outcomes with current approaches to digital PA interventions.

Methods: This review identified and summarized studies on digitally supported interventions for promoting PA in preschool-aged children. We generated 3 lists of relevant search terms that included technology-related terms, PA-related terms, and weight-related terms. The search included Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, and Daily, Ovid EMBASE, Ovid Cochrane Central Register of Controlled Trials, Ovid Cochrane Database of Systematic Reviews, and Scopus. Study selection was led by a single author and verified by a second; the same 2 authors assessed study quality using a standardized tool, and 3 authors completed data extraction on PA outcomes, theory application, and implementation outcomes.

Results: In total, 601 studies were identified; 8 met the inclusion criteria. For study quality, only 2 studies received an overall rating of strong quality and low risk of bias. All but 1 study had a small sample size (<100). Positive and significant changes in child PA outcomes were reported in only 2 studies with weak overall quality, both of which used child-directed approaches. In total, 5 studies applied a behavioral theory for designing the intervention; no patterns of effectiveness were identified based on the application of theory. Finally, no studies reported on the implementation outcomes of adoption, cost, penetration, or sustainability; 1 study did not assess any implementation outcomes, and no single study reported on more than 2 implementation outcomes. Studies measured the implementation outcome of acceptability most frequently (n=4), and researchers assessed fidelity in 3 studies.

Conclusions: The interventions with a significant effect on PA used child-centered activities; parent-directed digital interventions alone were ineffective for improving PA. Future research with rigorous designs, monitoring of implementation outcomes, and testing of the contributions of digital components will advance understanding of the effectiveness of digital interventions for increasing PA in children.

(J Med Internet Res 2022;24(1):e28230) doi:10.2196/28230

KEYWORDS

physical activity; preschool children; digital; technology; intervention

Introduction

Pediatric obesity is a major global health challenge jeopardizing development and well-being even beyond childhood [1-3]. Physical inactivity and sedentary lifestyle behaviors can contribute to childhood obesity. Regular physical activity (PA) in early life promotes healthy growth and development, improves children's' cardiovascular fitness, and promotes better motor and cognitive skill development [4]. Further, active children are less prone to develop chronic diseases later in life including cardiovascular diseases, type II diabetes mellitus, and obesity, as well as psychiatric, psychological, and psychosocial disorders when compared to inactive children [5-7].

Guidelines implemented in the United States for PA in preschool-aged children (3 to 5 years) encourage children to be active throughout the day by engaging in a variety of activities [8]. Although no specific amount is recommended, a reasonable target of 3 hours of movement—with 60 minutes of moderate to vigorous physical activity (MVPA)— is in line with the guidelines from the American Medical Association and the World Health Association [8,9]. Unfortunately, a significant proportion of preschool-aged children do not meet this target (less than 50% across studies) [7]. Several interventions designed to increase PA or reduce sedentary behaviors (eg, limit screen time) have resulted in inconsistent findings [10-12]. According to a meta-analysis of intervention studies with preschool-aged children, only small to moderate effects have been observed for improving PA, suggesting room for improvement in achieving the desired outcomes [13]. Effective PA-promoting interventions targeting preschool children are needed.

The use of newer technologies and digital platforms to mitigate sedentary behaviors and foster behaviors that increase children's PA may be a promising approach. Digital interventions have become more widely available globally, and health promotion through these platforms has become more accessible, easier to use, and more acceptable to families [14,15]. Indeed, the use of smartphones, websites, and text messaging offer relatively inexpensive and easy solutions to support or replace traditional face-to-face methods [16-19]. Studies with older children, aged 8 to 12 years, have used active video games (ie, exergaming) and digital applications to encourage PA participation [15,20]. Strategies that are child-centered, such as gamification including games with active plots and self-monitoring, have been used in digital applications [18,20]. These studies show that technologies merging PA and learning are available and may be helpful in promoting PA [21]. As the use of digital devices by young children has become common [22], digital platforms may hold significant promise for delivering PA interventions to preschoolers [3,23].

The relatively recent emergence of digital interventions and their potential to create equitable access to PA support also suggests the need to understand factors related to intervention success and failure. The application of behavioral theory for intervention design and implementation outcomes comprises at least 2 such factors. First, the use of theory allows intervention designers to be explicit with the targets for behavior change and select behavior change techniques that affect the theory-informed levers of change [24]. Second, measuring implementation outcomes in the delivery of digital interventions allows researchers to determine (1) if the observed effects (or the lack thereof) were attributable to implementation factors and (2) if the processes or characteristics of the interventions are desirable for larger-scale delivery [25]. Measuring implementation outcomes acknowledges that a potentially effective intervention can be implemented poorly (ie, implementation failure) or an intervention can be ineffective for a new setting (ie, intervention failure) [25]. Thus, assessing theoretical application and implementation outcomes are important for understanding current approaches to digital PA interventions.

There is considerable interest for expanding digital offerings to promote PA in young children, and studies are emerging that leverage digital interventions directed at children and their families. Thus, a review of existing studies to identify common or discriminating features of prior digital interventions contributing to improvements in child PA (or lack thereof) as well as factors associated with successful implementation of digital interventions for activity promotion is warranted. The aim of this systematic review was three-pronged: (1) to determine the quality of studies using digital PA intervention strategies with preschool-aged children; (2) to assess the efficacy of digital interventions and approaches designed to improve PA in preschool-aged children; and (3) to examine theoretical application and implementation outcomes with current approaches to digital PA interventions.

Methods

The systematic review followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-analyses) reporting guidelines [26].

Data Sources and Search Strategy

A comprehensive search of English language databases was conducted from each database's inception to January 24, 2020. The databases included Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations, and Daily, Ovid EMBASE, Ovid Cochrane Central Register of Controlled Trials, Ovid Cochrane Database of Systematic Reviews, and Scopus. The search strategy was designed and executed by an experienced librarian with input from the study's first and second authors (TS and AP). Controlled vocabulary supplemented with keywords was used to search for studies on the efficacy and acceptability of mobile-based, web-based, or other latest technology-supported interventions for promoting PA in preschool-aged children. We supplemented our professional librarian search using the same terms in Google Scholar and by manually searching and reviewing the reference lists of all included articles. The actual strategy listing all search terms

used and how they were combined is available in Multimedia Appendix 1.

Eligibility Criteria

Inclusion Criteria

The study designs considered in this review included randomized clinical and controlled trials, and quasi-experimental trials. Studies focusing on promoting PA in typically developing (without systemic, physical, and mental disorders) preschool-aged children (aged 3 to 5 years) were included. Interventions had to engage children or their parents using digitally based modalities to promote PA in children. Studies had to have PA-related outcomes as either the primary or secondary outcome and had to be published in English.

Exclusion Criteria

Studies were excluded if the preschool children were not included in the intervention or outcome assessment, if the focus was on health behaviors or conditions other than PA, or if children were not typically developing. Studies were excluded if digitally based platforms were not used as part of the intervention.

Outcome Variables

The primary outcome variable was PA, including subjectively and objectively determined PA levels (ie, self-reported or observed). PA outcomes included total PA, light physical activity (LPA), MVPA, percentage of time in LPA or MVPA (LPA% or MVPA%), energy expenditure, and steps. When available, weight-related outcomes were extracted as the secondary outcome including, but not limited to, BMI (kg/m^2), BMI Z-scores (BMIz), body fat (kg), body fat percentage, and waist circumference (cm). Other secondary outcome variables were implementation outcomes based on the taxonomy and definitions for studies reporting on implementation outcomes [25].

Selection of Articles

The first reviewer (AP) identified duplicates from the searches and screened the titles of the articles to shortlist target articles for review. The second reviewer (NZ) verified the first reviewer's decisions. The first and second reviewers independently screened the abstracts of the target articles and created a second shortlist. Thereafter, both reviewers read the full texts of the articles independently to evaluate them for inclusion in the final analysis. Discrepancies between the first and second reviewers were discussed and resolved by consensus. When a decision could not be reached, the coauthors reviewed the full texts to make the final decision.

Assessment of Study Quality and Risk of Bias

Study quality and risk of bias were assessed by 2 independent reviewers (AP and NZ) using the National Collaborating Centre for Methods and Tools Quality assessment tool for quantitative studies [27]. This assessment tool includes ratings (weak, moderate, and strong) for 6 components: (1) selection bias, (2) study design, (3) confounders, (4) blinding, (5) data collection methods, and (6) withdrawals and dropouts. Studies were rated as weak overall if 2 or more components were rated as weak, moderate if only 1 component was rated as weak, and strong if no components were rated as weak. Discrepancies were discussed and resolved by consensus between reviewers to ensure high agreement for ratings. Interrater reliability for individual component ratings was determined by computing the percentage of agreement and the Cohen κ.

Data Extraction and Synthesis for Study Characteristics, Intervention Efficacy, Use of Theory, and Implementation Outcomes

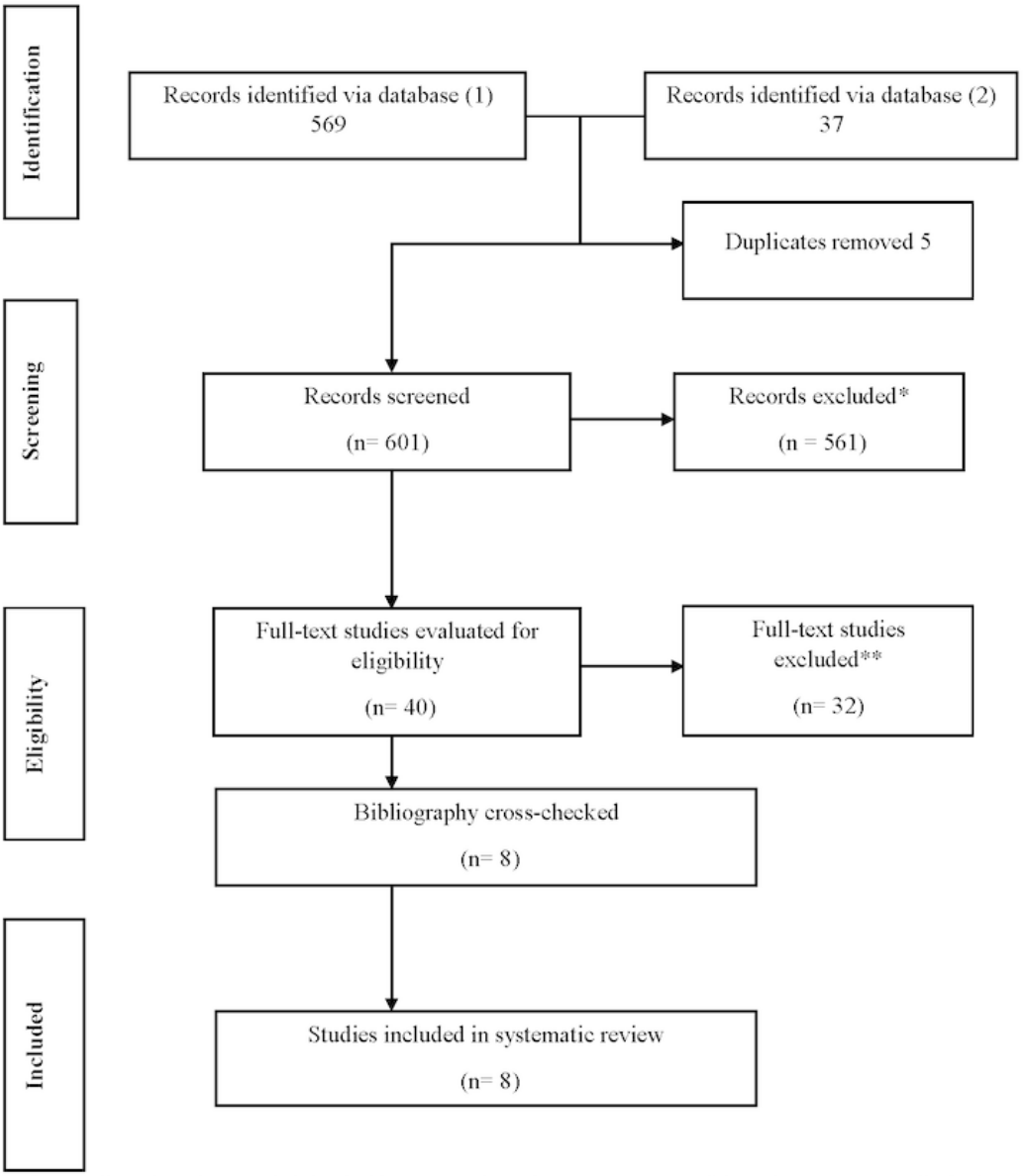
For primary extraction, each retrieved full-text article was evaluated systematically by 3 reviewers (NZ, AP, and TS) according to the following criteria: (1) design, (2) population, (3) intervention characteristics (modality, ie, how the intervention was delivered, exposure or dose, and content), (4) measures, and (5) analyses and results. As with data extraction, differences were resolved by mutual agreement; all coauthors reviewed the full texts to make the final decision when a decision could not be reached by consensus. TS extracted the theory used in each study, the content of each intervention, and secondary data on implementation, including implementation outcomes, implementation measures, data sources, and results.

Results

Selection of Articles

Figure 1 illustrates the search and selection process for articles. A total of 601 articles were identified initially. After removing duplicates and irrelevant studies, titles and abstracts of the remaining articles were further screened and identified as potentially meeting the inclusion criteria. Following a thorough assessment of the full-text articles, 8 studies fully met the inclusion criteria and were included in this review. The key reasons for excluding articles included ineligible age, special populations, no use of technology, no measure of PA, and non-English articles. A high interrater agreement (> 95%) for inclusion of the articles in this review was obtained between the authors (AP and NZ).

Figure 1. Flow chart for study inclusion.



*Reasons for study exclusion included ineligible age, special populations, no measures of physical activity, no use of technology.
**Reasons for study exclusion included ineligible age, special populations, no measures of physical activity, no use of technology, interventions that did not engage children either "and" or "or" their parents using digital-based modalities to promote physical activity in children, studies that did not report physical activity-related outcomes
Database (1): EBM Reviews - Cochrane Central Register of Controlled Trials December 2019, EBM Reviews - Cochrane Database of Systematic Reviews 2005 to January 21, 2020, Embase 1974 to January 23, 2020, Ovid MEDLINE(R) and Epub Ahead of Print, In-Process & Other Non-Indexed Citations and Daily 1946 to January 23, 2020
Database (2): Scopus- from database's inception to January 24th, 2020

Assessment of Study Quality and Risk of Bias

Following the ratings of the 6 steps in the National Collaborating Centre for Methods and Quality assessment tool for quantitative studies [27], the design quality and risk of bias for each study are presented in Multimedia Appendix 1. For the initial ratings, the agreement was 93%, and the Cohen κ=0.63. After resolving discrepancies, 2 studies received an overall rating of strong quality and low risk of bias [28,29]; 3 studies received an overall rating of moderate quality and medium risk of bias [30-32]; and 3 studies received an overall rating of weak quality and high risk of bias [33-35]. The most common issues with the study quality and risk of bias were related to selection bias and blinding. More specifically, all studies, apart from 1[31], failed to report the proportion of participants who agreed to participate and were thus unlikely to have a representative sample. Moreover, none of the studies was rated as strong on blinding, as most of the outcome assessors were aware of the research question and the exposure status, a common concern for behavioral intervention studies. Other issues included no reporting of the completion rate, unreliable data collection tools, and failure to control all relevant confounders.

Study Characteristics

Multimedia Appendix 2 summarizes details of the 8 included studies, ranked by quality. The exposure in most studies was to a technology-based behavior change program, most of which involved PA and nutrition components. All studies were

published within the last 6 years (2015-2019) and were conducted in Western countries, with 6 studies conducted in the United States [30-35], 1 in Australia [28], and 1 in Sweden [29]. Of the 8 studies, 5 primarily targeted White samples (41%-100% White) [28,29,31,33,35]; 2 studies targeted samples that were mostly Asian (56% and 100%) [30,32], and 1 study targeted a sample that was mainly Black (42%) [34]. Among these studies, 6 were randomized controlled trials [28-33], and 2 were quasi-experimental trials [34,35]. Of the 8 studies, 3 primarily targeted preschoolers [30,33,35]; 1 had components directed at children and parents [34]; and the remaining 4 studies were aimed at parents [28,29,31,32]. Of the 8 included studies, 2 were feasibility studies with small sample sizes [32,34]. Sample sizes varied across studies, ranging from 32 to 315 participants; 7 studies included <100 participants in total. Control groups or conditions included usual care (eg, Head Start and recess as usual) [30,34,35], free play for children [33], educational resources [28,29,32], and knowledge-based lifestyle programs [31].

Technology modalities in these studies included exergaming (n=3) [30,33,35], social media (Facebook, n=1) with education and social support for parents [35], websites with education modules for parents (n=1) [31], a mobile app with information and strategies for parents (n=1) [29], a preloaded tablet with educational modules for parents (n=1) [32], and a combination of delivery modes including educational and motivational content for parents (eg, emails, social media, online video, telephone calls, and text messages, n=1) [28]. Most studies used a digital platform to deliver health information and education to parents; only the 3 studies evaluating exergaming interventions provided direct PA opportunities to children via a digital modality.

Intervention lengths ranged from 6 weeks to 6 months. Intervention intensity ranged from no required PA sessions to daily 30-minute sessions for 12 weeks. Of the 8 studies, 7 focused on short-term outcomes (≤2 months), and 1 focused on medium-term outcomes (3 to 6 months) [31].

Outcome assessment included objective and subjective measures. In total, 7 studies used objective assessments of PA, including accelerometers and pedometers to capture participants' (children or parents) PA levels [28-30,33-35]. PA was assessed subjectively in 2 studies [31,32] using parent-reported questionnaires to interpret PA behaviors. All studies assessing weight-related outcomes used objective measures [28-30,32,34]. In addition to PA outcomes, several studies included parent-centered questionnaires related to other targeted behaviors (eg, sedentary behavior, dietary intake, sleep, screen time, feeding practices, parent role modeling, self-efficacy, beverage consumption, and eating behaviors) [28,29,31,32,34]. Child-centered questionnaires were related to enjoyment of movement, perceived competence, and motor skill competence [33,35]. Other objective measures included assessments of cardiovascular fitness, gross motor development, and cognitive flexibility of the children [33,35].

Intervention Efficacy

PA Outcomes

Positive changes in the PA outcomes of children were reported in only 2 studies [33,35]. Fu et al [33] documented an increase of 887 steps via a pedometer among children receiving the intervention compared to the control group, whereas Gao et al [35] demonstrated significant increases in the MVPA time via an accelerometer at the end of the intervention (ie, an increase of 4 minutes per day for the treatment group and a decrease of 2 minutes per day for the control group). Both studies reflect short-term, school-based, child-directed approaches with the intervention centered on providing direct opportunities for PA through exergaming (eg, GoNoodle and Wii Fit).

Weight-Related Outcomes

None of the 5 studies examining children's weight-related outcomes demonstrated significant effects [28-30,32].

Other Health Outcomes

Studies produced mixed results on other health-related metrics. In addition to favoring PA outcomes, 2 child-centered (exergaming) interventions reported positive effects on the children's perceived motor competence [33], gross motor development [33], and cognitive flexibility [30]. In addition, parent-directed interventions produced positive effects for improving self-efficacy [28,31,32] with an exception [34]. No intervention aiming to decrease screen time achieved the desired outcome [28,31,34]. Other outcomes demonstrated no clear patterns of effects.

Use of Theory and Implementation Outcomes

Multimedia Appendix 3 presents the theory applied by each study and the implementation outcomes assessed. The most commonly used behavioral theory was social cognitive theory (n=3) [28,29,31]. Other theories included the actor-partner interdependence model [34] and the information-motivation-behavior model [32]. Further, 3 studies did not report the explicit use of theories [30,33,35]. No clear patterns of effectiveness were identified based on the application of theory.

No studies reported on the implementation outcomes of adoption, cost, penetration, or sustainability; 1 study did not assess any implementation outcomes [33], and no single study reported on more than 2 implementation outcomes. Studies measured the implementation outcome of acceptability most frequently (n=4), with measures including self-developed surveys [28,34], interviews with participants [32,34], and data from the technology platforms (eg, frequency and duration of use) [31]. The 2 studies with ratings of strong overall quality only examined the implementation outcome of acceptability [28,29]. Researchers assessed fidelity in 3 studies, with 2 monitoring fidelity through observing intervention sessions and completing self-developed checklists [30,35], and 1 study assessed fidelity through a user-completed survey at the end of the intervention [31]. Researchers captured indicators of feasibility and appropriateness in 2 studies. Appropriateness assessments were paired with questions about acceptability in both studies; 1 study focused on the appropriateness of specific

intervention aspects (eg, length) through a user-completed survey at the end of the intervention [28], and the other assessed cultural appropriateness in focus groups (preintervention) and interviews (postintervention) [32]. The lack of data on implementation outcomes precludes the ability to link efficacy and implementation outcomes.

Discussion

Principal Findings

Prior reviews have shown the value of digital platforms for improving PA-related outcomes in children aged 6 to 12 years [36], adolescents [37], adults [38,39], and older adults [40]; this study sought to determine if similar documented effects exist for children aged 3 to 5 years. Of the 8 studies identified, all were published in the last 6 years with 2 studies showing positive effects on PA in children and only 2 demonstrating a strong quality and low risk of bias (neither of which showed PA effects). Studies measuring implementation outcomes most frequently assessed indicators of acceptability; however, implementation outcomes were not a prominent focus of the studies included. In addition, although most studies applied a theoretical framework, no clear patterns were noted based on use of theory. These findings illustrate the early stages of exploration in this scientific area and the opportunity to conduct future studies that are more rigorous in their design. With such a low number of studies showing changes in activity for children, our ability to examine shared features and patterns was limited.

Based on the limited number of studies showing improved PA in preschool children, common patterns in the effective studies are tentatively noted. Notably, the studies showing effective interventions used child-centered activities (exergaming) in schools; the exergames used in these studies were commercially available (ie, no new intervention or technology was developed). A narrative review noted the strengths of exergaming, including its adaptability, customizability, and scalability for reaching children and adolescents; weaknesses include the lack of sustained engagement over time and costs of development [20]. As technology evolves, it is imperative to understand the attributes of effective child-centered digital activities, such as exergaming, and how other digital platforms (mobile apps, virtual reality, and web-based games) can replicate these outcomes.

Given the limited number of studies in this nascent area of research, this review is more useful to highlight gaps for future work and identify the features of ineffective studies. First, studies aimed at parent education without direct intervention for children did not show desired increases in PA. This finding contradicts a recent review on technology-based and parent-targeted tools for improving nutrition outcomes in children that found positive effects in 10 out of 11 studies [41]. This difference suggests that parents' gatekeeping behaviors may have a more direct influence on children's dietary intake than movement habits; engaging parents and children in PA may better support change in this area. Debates about the value of engaging parents versus parent-child dyads or children alone is not new to intervention research. Understanding the effects of the bidirectional relationship between parents and children, particularly the mutual influence of parenting behaviors and children's outcomes in early childhood, has the potential to enhance the development of digital interventions and their positioning to target audiences [42]. None of the interventions reviewed herein engaged children and parents through the same intervention for mutual increase in PA; this remains an area for future research. Recent formative research suggests that parents are open to the use of digital applications to support such an approach [3].

A second feature of the ineffective studies was the use of subjective measures. Although only 2 studies used subjective measures of PA, they did not show changes after intervention. This finding may suggest the superiority of objective measures of PA for identifying intervention-induced changes, consistent with prior reviews that advocate for standardized accelerometer evaluation in PA interventions [43] with young children. Objective measures may also be more suitable to capture the sporadic short bursts of movement, which tend to typify PA in early childhood and can be missed with subjective assessments. Nevertheless, objective measures do not capture PA and parenting styles, environmental conditions or stimuli, or PA behaviors and patterns of other family members, all of which can impact PA in young children.

Lastly, studies that addressed multiple outcome targets (ie, PA and nutrition) and modalities (websites + face-to-face interventions) did not have clear advantages. This is in contrast to a review of studies regarding the effects of mobile apps on health outcomes across age groups, which showed that multimodality studies had greater effects [44]. The lack of this documented advantage in our review may be due to the limited number of studies and the lack of high-quality studies without bias. Future intervention work should consider exploring the impact of individual intervention components versus the intervention as a whole. This would provide insights into how digital modalities are influencing or driving intervention outcomes.

The most significant findings of this review lie in highlighting opportunities for further work in this area related to the quality of the studies, intervention strategies, and inclusion of implementation outcomes. Except for 1 study, the rest had samples sizes less than 100; future work can expand the size and diversity of samples to understand the different settings and circumstances in which technology will work to increase PA in children (ie, evaluate mediators and moderators). Only 1 study assessed outcomes after 6 months postintervention. Future work should extend the time of follow-up used in prior studies, which have noted that technology-based intervention effects may not be long-lasting. Such a study is in progress that will examine the effects of a mobile app on children's activity's levels over 24 weeks [45].

Regarding the use of digital platforms as an intervention strategy for young children, prior studies have raised concerns that digital platforms may undermine traditional forms of exercise and contribute to excessive screen time in children [46]. Future work must examine if this is true. Conversely, young children who are active in using technology may increase activity in their

daily routines (beyond technology time); this means technology time displaces other screen time rather than increasing it. These outcomes were not assessed in the studies presented in this report. Digital interventions have the potential to reach children and families with PA interventions in areas where structured, face-to-face programs are limited [23].

Combined with the inconsistent collection of implementation outcomes, studies failing to find effects may be unable to distinguish failure of the intervention from failure of the study design or from failure of implementation. Work beyond assessing acceptability (after feasibility is established) will be needed to provide better understanding of the implementation process for using technology to target young children for improving PA. As digitally based interventions are integrated into existing systems (eg, childcare, schools, and homes), the implementation outcomes missing in these studies will become particularly important (ie, adoption, cost, penetration, and sustainment). Further, future interventions could benefit from assessing the adoption rates by participant characteristics, costs of intervention delivery to inform future scalability, and sustained use of digital strategies beyond research study contexts [37].

The key limitation of this report is the lack of conclusive statements about digitally based PA interventions due to the emerging nature of the field. This is reflected in the low number and limited quality of prior studies in this area to date. Further, our review included only peer-reviewed full-text and English language publications; other unpublished and non-English research papers may exist on this topic. These factors limited the conclusions that could be drawn from this review. The strengths of our work include the systematic approach, use of the most updated guidelines on completion of systematic reviews; assessment of study quality, bias, and strength of the evidence; and high agreement levels between the reviewers.

Conclusions

Across a range of physical health, psychological health, and cognitive development aspects, PA has been associated with positive outcomes for young children. Effective and scalable intervention methods are needed to help children achieve the recommended levels of PA and movement during their preschool years. This review suggests that interventions involving child-centered approaches may be the most promising for increasing PA in children. Specifically, commercial products packaged for delivery as PA interventions in schools used in 2 studies had significant effects on PA according to this review. This indicates opportunities to develop new digitally based interventions that are designed considering the needs of young users and their families. Although this review demonstrates that digital platforms are promising to a certain extent for achieving increased PA in children, there are numerous gaps in the current evidence. Further research using rigorous designs to achieve high study quality and minimize bias, monitoring implementation outcomes, and distinguishing the contributions of digital components from other intervention components will advance the understanding of the effectiveness of interventions and their potential to be implemented more broadly (versus those with characteristics limiting implementability). Additionally, future research focus should be on rigorous study designs involving diverse populations engaged in interventions delivered to the family and children that include objectively measured PA levels as primary or secondary outcomes.

Acknowledgments

We would like to thank the librarians at the Mayo Clinic (Rochester, MN) and University of Arkansas for Medical Sciences for their assistance with designing and executing the literature search. This work is supported by the United States Department of Agriculture, grant number USDA - ARS Project 6026 - 51000 - 012 - 06S. The funder provided approval of the study design.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Quality assessment of included studies.
[DOCX File , 24 KB - jmir_v24i1e28230_app1.docx]

Multimedia Appendix 2
Characteristics of included studies by global quality rating.
[DOCX File , 43 KB - jmir_v24i1e28230_app2.docx]

Multimedia Appendix 3
Implementation characteristics and outcomes of included studies by Global Quality Rating.
[DOCX File , 29 KB - jmir_v24i1e28230_app3.docx]

References

1. Bahreynian M, Qorbani M, Khaniabadi BM, Motlagh ME, Safari O, Asayesh H, et al. Association between obesity and parental weight status in children and adolescents. J Clin Res Pediatr Endocrinol 2017 Jun;9(2):111-117 [FREE Full text] [doi: 10.4274/jcrpe.3790] [Medline: 28008863]
2. A Hamid MS, Sazlina SG. Interventions for obesity among schoolchildren: a systematic review and meta-analyses. PLoS One 2019 Jan;14(1):e0209746 [FREE Full text] [doi: 10.1371/journal.pone.0209746] [Medline: 30625165]
3. Ek A, Sandborg J, Delisle Nyström C, Lindqvist A, Rutberg S, Löf M. Physical activity and mobile phone apps in the preschool age: perceptions of teachers and parents. JMIR Mhealth Uhealth 2019 Apr;7(4):e12512 [FREE Full text] [doi: 10.2196/12512] [Medline: 30994465]
4. Zeng N, Ayyub M, Sun H, Wen X, Xiang P, Gao Z. Effects of physical activity on motor skills and cognitive development in early childhood: a systematic review. Biomed Res Int 2017 Dec;2017:2760716 [FREE Full text] [doi: 10.1155/2017/2760716] [Medline: 29387718]
5. Ozemek C, Lavie CJ, Rognmo. Global physical activity levels-need for intervention. Prog Cardiovasc Dis 2019 Mar;62(2):102-107. [doi: 10.1016/j.pcad.2019.02.004] [Medline: 30802461]
6. Booth FW, Roberts CK, Thyfault JP, Ruegsegger GN, Toedebusch RG. Role of inactivity in chronic diseases: evolutionary insight and pathophysiological mechanisms. Physiol Rev 2017 Oct;97(4):1351-1402. [doi: 10.1152/physrev.00019.2016] [Medline: 28814614]
7. Tucker P. The physical activity levels of preschool-aged children: a systematic review. Early Child Res Q 2008 Oct;23(4):547-558. [doi: 10.1016/j.ecresq.2008.08.005]
8. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The physical activity guidelines for Americans. JAMA 2018 Nov;320(19):2020-2028. [doi: 10.1001/jama.2018.14854] [Medline: 30418471]
9. Guidelines on physical activity, sedentary behaviour, and sleep for children under 5 years of age. Wolrd Health Organization. 2019. URL: https://apps.who.int/iris/handle/10665/311664 [accessed 2020-08-10]
10. Morgan EH, Schoonees A, Sriram U, Faure M, Seguin-Fowler RA. Caregiver involvement in interventions for improving children's dietary intake and physical activity behaviors. Cochrane Database Syst Rev 2020 Jan;1:CD012547 [FREE Full text] [doi: 10.1002/14651858.CD012547.pub2] [Medline: 31902132]
11. Johnstone A, Hughes AR, Martin A, Reilly JJ. Utilising active play interventions to promote physical activity and improve fundamental movement skills in children: a systematic review and meta-analysis. BMC Public Health 2018 Jun;18(1):789 [FREE Full text] [doi: 10.1186/s12889-018-5687-z] [Medline: 29940923]
12. Lee AM, Chavez S, Bian J, Thompson LA, Gurka MJ, Williamson VG, et al. Efficacy and effectiveness of mobile health technologies for facilitating physical activity in adolescents: scoping review. JMIR Mhealth Uhealth 2019 Feb;7(2):e11847 [FREE Full text] [doi: 10.2196/11847] [Medline: 30747716]
13. Gordon ES, Tucker P, Burke SM, Carron AV. Effectiveness of physical activity interventions for preschoolers: a meta-analysis. Res Q Exerc Sport 2013 Sep;84(3):287-294. [doi: 10.1080/02701367.2013.813894]
14. Tate EB, Spruijt-Metz D, O'Reilly G, Jordan-Marsh M, Gotsis M, Pentz MA, et al. mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. Transl Behav Med 2013 Dec;3(4):406-415 [FREE Full text] [doi: 10.1007/s13142-013-0222-3] [Medline: 24294329]
15. Kozak AT, Buscemi J, Hawkins MAW, Wang ML, Breland JY, Ross KM, et al. Technology-based interventions for weight management: current randomized controlled trial evidence and future directions. J Behav Med 2017 Feb;40(1):99-111. [doi: 10.1007/s10865-016-9805-z] [Medline: 27783259]
16. Staiano A, Broyles S, Katzmarzyk P. School term vs. school holiday: associations with children's physical activity, screen-time, diet and sleep. Int J Environ Res Public Health 2015 Jul;12(8):8861-8870. [doi: 10.3390/ijerph120808861]
17. Murray E, Hekler EB, Andersson G, Collins LM, Doherty A, Hollis C, et al. Evaluating digital health interventions: key questions and approaches. Am J Prev Med 2016 Nov;51(5):843-851. [doi: 10.1016/j.amepre.2016.06.008] [Medline: 27745684]
18. Johnson D, Deterding S, Kuhn K, Staneva A, Stoyanov S, Hides L. Gamification for health and wellbeing: a systematic review of the literature. Internet Interv 2016 Nov;6:89-106 [FREE Full text] [doi: 10.1016/j.invent.2016.10.002] [Medline: 30135818]
19. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
20. Benzing V, Schmidt M. Exergaming for children and adolescents: strengths, weaknesses, opportunities and threats. J Clin Med 2018 Nov 08;7(11):422 [FREE Full text] [doi: 10.3390/jcm7110422] [Medline: 30413016]
21. Silverstone S, Teatum J. Technology: the problem or the solution to childhood obesity. Am J Bus Educ 2011:37-60 [FREE Full text] [doi: 10.19030/ajbe.v4i1.1272]
22. Domoff SE, Radesky JS, Harrison K, Riley H, Lumeng JC, Miller AL. A naturalistic study of child and family screen media and mobile device use. J Child Fam Stud 2019 Feb;28(2):401-410 [FREE Full text] [doi: 10.1007/s10826-018-1275-1] [Medline: 31105418]

23. McCloskey ML, Thompson DA, Chamberlin B, Clark L, Johnson SL, Bellows LL. Mobile device use among rural, low-income families and the feasibility of an app to encourage preschoolers' physical activity: qualitative study. JMIR Pediatr Parent 2018 Dec;1(2):e10858 [FREE Full text] [doi: 10.2196/10858] [Medline: 31518295]
24. Lippke S, Ziegelmann J. Theory-based health behavior change: developing, testing, and applying theories for evidence-based interventions. Appl Psychol 2008 Oct;57(4):698-716. [doi: 10.1111/j.1464-0597.2008.00339.x]
25. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
26. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 2009 Oct;62(10):e1-e34 [FREE Full text] [doi: 10.1016/j.jclinepi.2009.06.006] [Medline: 19631507]
27. Thomas B, Ciliska D, Dobbins M, Micucci S. A process for systematically reviewing the literature: providing the research evidence for public health nursing interventions. Worldviews Evid Based Nurs 2004 Sep;1(3):176-184. [doi: 10.1111/j.1524-475X.2004.04006.x]
28. Hammersley ML, Okely AD, Batterham MJ, Jones RA. An internet-based childhood obesity prevention program (Time2bHealthy) for parents of preschool-aged children: randomized controlled trial. J Med Internet Res 2019 Feb;21(2):e11964 [FREE Full text] [doi: 10.2196/11964] [Medline: 30735139]
29. Nyström CD, Sandin S, Henriksson P, Henriksson H, Trolle-Lagerros Y, Larsson C, et al. Mobile-based intervention intended to stop obesity in preschool-aged children: the MINISTOP randomized controlled trial. Am J Clin Nutr 2017 Jun;105(6):1327-1335. [doi: 10.3945/ajcn.116.150995] [Medline: 28446496]
30. Gao Z, Lee JE, Zeng N, Pope ZC, Zhang Y, Li X. Home-based exergaming on preschoolers' energy expenditure, cardiovascular fitness, body mass index and cognitive flexibility: a randomized controlled trial. J Clin Med 2019 Oct;8(10):1745 [FREE Full text] [doi: 10.3390/jcm8101745] [Medline: 31640158]
31. Knowlden AP, Sharma M, Cottrell RR, Wilson BRA, Johnson ML. Impact evaluation of enabling mothers to prevent pediatric obesity through web-based education and reciprocal determinism (EMPOWER) randomized control trial. Health Educ Behav 2015 Apr;42(2):171-184. [doi: 10.1177/1090198114547816] [Medline: 25161168]
32. Sun A, Cheng J, Bui Q, Liang Y, Ng T, Chen J. Home-based and technology-centered childhood obesity prevention for Chinese mothers with preschool-aged children. J Transcult Nurs 2017 Nov;28(6):616-624. [doi: 10.1177/1043659617719139] [Medline: 28826348]
33. Fu Y, Burns RD, Constantino N, Zhang P. Differences in step counts, motor competence, and enjoyment between an exergaming group and a non-exergaming group. Games Health J 2018 Oct;7(5):335-340. [doi: 10.1089/g4h.2017.0188] [Medline: 29989430]
34. Ling J, Robbins LB, Zhang N, Kerver JM, Lyons H, Wieber N, et al. Using Facebook in a healthy lifestyle intervention: feasibility and preliminary efficacy. West J Nurs Res 2018 Dec;40(12):1818-1842. [doi: 10.1177/0193945918756870] [Medline: 29421989]
35. Gao Z, Zeng N, Pope ZC, Wang R, Yu F. Effects of exergaming on motor skill competence, perceived competence, and physical activity in preschool children. J Sport Health Sci 2019 Mar;8(2):106-113. [doi: 10.1016/j.jshs.2018.12.001]
36. Lau PWC, Lau EY, Wong DP, Ransdell L. A systematic review of information and communication technology-based interventions for promoting physical activity behavior change in children and adolescents. J Med Internet Res 2011 Jul;13(3):e48 [FREE Full text] [doi: 10.2196/jmir.1533] [Medline: 21749967]
37. Rose T, Barker M, Maria JC, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
38. Stephenson A, McDonough SM, Murphy MH, Nugent CD, Mair JL. Using computer, mobile and wearable technology enhanced interventions to reduce sedentary behaviour: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2017 Aug;14(1):105 [FREE Full text] [doi: 10.1186/s12966-017-0561-4] [Medline: 28800736]
39. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-based wearable activity trackers increase physical activity participation: systematic review and meta-analysis. JMIR Mhealth Uhealth 2019 Apr;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
40. Yerrakalva D, Yerrakalva D, Hajna S, Griffin S. Effects of mobile health app interventions on sedentary time, physical activity, and fitness in older adults: systematic review and meta-analysis. J Med Internet Res 2019 Nov;21(11):e14343 [FREE Full text] [doi: 10.2196/14343] [Medline: 31778121]
41. Zarnowiecki D, Mauch CE, Middleton G, Matwiejczyk L, Watson WL, Dibbs J, et al. A systematic evaluation of digital nutrition promotion websites and apps for supporting parents to influence children's nutrition. Int J Behav Nutr Phys Act 2020 Feb;17(1):17 [FREE Full text] [doi: 10.1186/s12966-020-0915-1] [Medline: 32041640]
42. Paschall KW, Mastergeorge AM. A review of 25 years of research in bidirectionality in parent–child relationships. Int J Behav Dev 2016 Jul;40(5):442-451. [doi: 10.1177/0165025415607379]

43. Carson V, Lee E, Hewitt L, Jennings C, Hunter S, Kuzik N, et al. Systematic review of the relationships between physical activity and health indicators in the early years (0-4 years). BMC Public Health 2017 Nov;17(Suppl 5):854 [FREE Full text] [doi: 10.1186/s12889-017-4860-0] [Medline: 29219090]
44. Schoeppe S, Alley S, Van Lippevelde W, Bray NA, Williams SL, Duncan MJ, et al. Efficacy of interventions that use apps to improve diet, physical activity and sedentary behaviour: a systematic review. Int J Behav Nutr Phys Act 2016 Dec;13(1):127 [FREE Full text] [doi: 10.1186/s12966-016-0454-y] [Medline: 27927218]
45. Webster EK, Kracht CL, Newton RL, Beyl RA, Staiano AE. Intervention to improve preschool children's fundamental motor skills: protocol for a parent-focused, mobile app-based comparative effectiveness trial. JMIR Res Protoc 2020 Oct;9(10):e19943 [FREE Full text] [doi: 10.2196/19943] [Medline: 33079066]
46. Gao Z, Lee JE. Emerging technology in promoting physical activity and health: challenges and opportunities. J Clin Med 2019 Nov;8(11):1830 [FREE Full text] [doi: 10.3390/jcm8111830] [Medline: 31683951]

Abbreviations

LPA: light physical activity
MVPA: moderate to vigorous physical activity
PA: physical activity

Edited by R Kukafka, G Eysenbach; submitted 25.02.21; peer-reviewed by R Kretschmann, C Miranda, S Subramaniam; comments to author 10.05.21; revised version received 27.05.21; accepted 05.10.21; published 11.01.22.

Please cite as:
Swindle T, Poosala AB, Zeng N, Børsheim E, Andres A, Bellows LL
Digital Intervention Strategies for Increasing Physical Activity Among Preschoolers: Systematic Review
J Med Internet Res 2022;24(1):e28230
URL: https://www.jmir.org/2022/1/e28230
doi:10.2196/28230
PMID:35014962

Review

Understanding the Nature of Metadata: Systematic Review

Hannes Ulrich[1,2], MSc; Ann-Kristin Kock-Schoppenhauer[1], MSc; Noemi Deppenwiese[3], MSc; Robert Gött[4], MSc; Jori Kern[5,6], BSc; Martin Lablans[5,6], PhD; Raphael W Majeed[7,8], MSc; Mark R Stöhr[7]; Jürgen Stausberg[9], MD; Julian Varghese[10], MSc, MD; Martin Dugas[11], MSc, MD; Josef Ingenerf[1,2], PhD

[1]IT Center for Clinical Research, University of Lübeck, Lübeck, Germany

[2]Institute of Medical Informatics, University of Lübeck, Lübeck, Germany

[3]Chair of Medical Informatics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany

[4]Department Epidemiology of Health Care and Community Health, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany

[5]Federated Information Systems, German Cancer Research Center, Heidelberg, Germany

[6]Complex Data Processing in Medical Informatics, University Medical Center Mannheim, Mannheim, Germany

[7]Universities of Giessen and Marburg Lung Center, German Center for Lung Research, Justus-Liebig-University, Giessen, Germany

[8]Institute of Medical Informatics, University Hospital RWTH Aachen, Aachen, Germany

[9]Institute of Medical Informatics, Biometry and Epidemiology, Faculty of Medicine, University of Duisburg-Essen, Essen, Germany

[10]Institute of Medical Informatics, University of Münster, Münster, Germany

[11]Institute of Medical Informatics, Heidelberg University Hospital, Heidelberg, Germany

Corresponding Author:
Hannes Ulrich, MSc
IT Center for Clinical Research
University of Lübeck
Ratzeburger Allee 160
Lübeck, 23564
Germany
Phone: 49 45131015607
Email: h.ulrich@uni-luebeck.de

Abstract

Background: Metadata are created to describe the corresponding data in a detailed and unambiguous way and is used for various applications in different research areas, for example, data identification and classification. However, a clear definition of metadata is crucial for further use. Unfortunately, extensive experience with the processing and management of metadata has shown that the term “metadata” and its use is not always unambiguous.

Objective: This study aimed to understand the definition of metadata and the challenges resulting from metadata reuse.

Methods: A systematic literature search was performed in this study following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for reporting on systematic reviews. Five research questions were identified to streamline the review process, addressing metadata characteristics, metadata standards, use cases, and problems encountered. This review was preceded by a harmonization process to achieve a general understanding of the terms used.

Results: The harmonization process resulted in a clear set of definitions for metadata processing focusing on data integration. The following literature review was conducted by 10 reviewers with different backgrounds and using the harmonized definitions. This study included 81 peer-reviewed papers from the last decade after applying various filtering steps to identify the most relevant papers. The 5 research questions could be answered, resulting in a broad overview of the standards, use cases, problems, and corresponding solutions for the application of metadata in different research areas.

Conclusions: Metadata can be a powerful tool for identifying, describing, and processing information, but its meaningful creation is costly and challenging. This review process uncovered many standards, use cases, problems, and solutions for dealing with metadata. The presented harmonized definitions and the new schema have the potential to improve the classification and generation of metadata by creating a shared understanding of metadata and its context.

(J Med Internet Res 2022;24(1):e25440) doi:10.2196/25440

KEYWORDS

metadata; metadata definition; systematic review; data integration; data identification; data classification

Introduction

Computer-aided medicine is revolutionizing health care and is creating treatment possibilities that are unimaginable without computer assistance: personalized medicine, improved diagnostics by artificial intelligence, and robot-assisted surgery. An immense amount of data fuels this digital revolution, and it is desperately needed for specialized procedures to be developed and optimized. This information is primarily created to document patient care for legal or financial purposes [1] and is often stored in silos [2], consequently making it hard to reach and impossible to reuse. Owing to the missing exchange, data formats will differ, creating data heterogeneity, which is a well-discussed issue in computer science [3]. Metadata can support the integration of heterogeneous data sources to achieve a valid and meaningful data fusion, enabling a comprehensible reuse of the stored medical information [4]. Metadata are created for a detailed and unique description of the corresponding data. It serves various use cases in different research areas, for example, data identification, classification, retrieval, and data set validation. The unambiguous and precise definition of metadata is crucial and is increasingly becoming a focus of active research. An important aspect of the research is the proposed findability, accessibility, interoperability, and reusability principles by Wilkinson et al [5], which are clear guidelines for the association of data and metadata. However, from current experiences, the definition of the term “metadata” is far from clear and very nonuniformly applied in everyday life. The problem is the variety of definitions, formats, standards, and contexts, which leads to a vague understanding of the actual metadata itself. The harmonization aspect, which was intended to be solved by using metadata, resulted in another form of heterogeneity instead of a solution for missing interoperability. It appears that domain experts providing clinical metadata and metadata experts have different definitions and boundaries of 2 central metadata concepts: the definition of the metadata itself and metadata composition (like matching, mapping, and transformation). To our knowledge, there exists no analysis on these concepts found in the current literature. To close this knowledge gap, we performed an expert review using the literature from the last decade. The review’s focus and the proposed research questions were driven by the issues and misunderstanding experiences on a daily basis in our intersectoral projects. Thus, a precisely defined harmonized understanding of the term “metadata” would therefore be indispensable for current and future developments in all aspects of data integration. To ensure a wide definition of metadata, various research fields (including social science, geography, and bibliography) were investigated for metadata applications, focusing on the described problems, provided solutions, and their transferability to the field of medical informatics.

Methods

Design

The systematic literature review performed in this study was done following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for reporting systematic reviews. A harmonization process preceded the review to gain a general understanding of the used terms.

Harmonization Process for Quality Assurance of the Review

This review was performed by 10 reviewers with expertise in medical informatics and technical and semantic interoperability [6] of medical metadata. All reviewers had different professional backgrounds: physicians, medical computer scientists with different technical expertise, and metadata curators. Initially, we recognized a missing general use of the technical terms in the field of metadata. Therefore, to guarantee a consistent understanding of the terms among the experts and to minimize the misinterpretation and misclassification during the analysis process, the actual review was preceded by a harmonization process resulting in a joint agreement on the definitions of metadata matching, mapping, and transformation. A questionnaire was created containing 5 questions and tasks concerning the scientific background in metadata, classifications of metadata processing, and their potential for automation and the definitions of metadata matching, mapping, and transformation.

Systematic Review

The PRISMA guidelines were applied in the systematic review as a de facto standard [7]. The process started by defining distinct and clear research questions that should be answered by the literature review. Daily work with metadata for clinical data integration has shown that there is no clear understanding of metadata and its potential applications by the users and experts. As an example, matching can be understood in various ways. Metadata matches to instance data [8] or to semantic attributes [9] or other metadata [10]. The general understanding is ambiguous. Therefore, our study aimed to explore to find an acceptable definition of metadata (Q1) and, with our operational focus on data integration, definitions for metadata processing (Q2) to enhance our daily operational tasks. In addition, we aimed to provide an overview of the variety of metadata standards used (Q3) and the generation of metadata in other research domains (Q4) to understand the issues involved and how they are solved (Q5). Thus, the focus questions were as follows:

Q1: How is the term “metadata” defined in different research fields?

Q2: How are the terms “metadata matching” and “metadata mapping” defined?

Q3: Which standards concerning metadata are in use?

Q4: How are metadata created in other research fields?

Q5: What are the current problems regarding the use of metadata, and which solutions are mentioned?

Data Sources and Search Criteria

The review and its results were based on extensive literature analysis; therefore, the selected literature was extremely important to the results. In this review, Scopus and Web of Science was used. The selection phase was 2-fold: in the first step, the very general keyword “metadata” was used to obtain a wide variety of publications. The search query was restricted to include only journal papers, conference proceedings, and book chapters from the last 10 years (2010-2019). About 11.6% (2453/21,161) of the resulting papers were randomly selected and then analyzed by title and abstract to identify papers within the scope of the research questions. Potential publications that were of uncertain use were included at this stage to prevent hasty exclusion. The keywords of suitable papers were used in the second step of the literature search for the full-text analysis. The papers of the second literature query were analyzed by titles and abstracts again to match the research questions for inclusion in the full-text analysis.

Review Process

Each of the 81 papers was reviewed by the first author and 2 randomly assigned reviewers, resulting in 3 independent interpretations per paper. To standardize the review process, a survey form with 8 questions was created: 6 questions corresponding to the research focus and 2 questions to gain additional information about the selected literature. The main questions focused on the metadata definitions (Q1), scoping metadata matching, mapping, and transformation (Q2), used standards (Q3), applied use cases (Q4), encountered problems, and the corresponding solutions (Q5). The additional questions covered the research field from which the paper originated and which type of metadata is described. For the categorization of the metadata types, a classification published by the National Information Standards Organization (NISO) [11] was used, which should help to classify metadata into the introduced categories better. This classification introduced 3 different types:

- descriptive metadata describe a resource for discovery and identification purposes,
- structural metadata describe the schema, data models, and reference data, and
- administrative metadata provide information about the management of a resource.

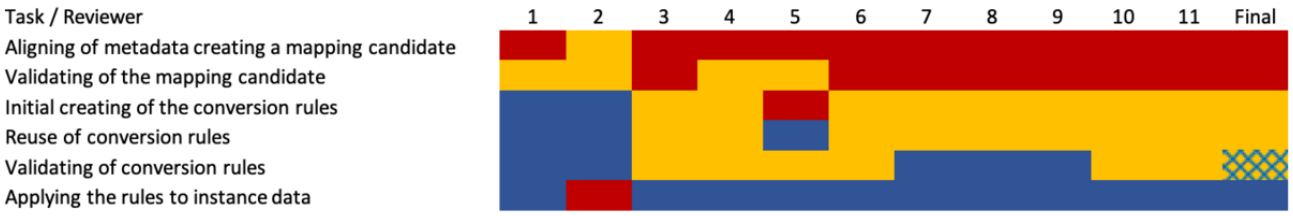
To illustrate the classification, consider this example: a book can be described using 3 different types of metadata. Author, title, and preface are examples for descriptive information, whereas the arrangement in chapters and page ordering is structural metadata. Information about the publication date and copyright information is classified as administrative metadata. The review process was open for 8 weeks. The results were gathered and analyzed by the first author and verified by the reviewers to produce a joint agreement on the final results. Both survey forms and the review results can be found in Multimedia Appendix 1 and Multimedia Appendix 2.

Results

Harmonized Definitions for Metadata Processing

Ten reviewers participated in the harmonization process. The reviewers categorized 6 metadata processing tasks concerning the use case of metadata-driven data integration as matching, mapping, or transformation. Furthermore, the reviewers assessed to which degree the metadata processing tasks can be automated. The results showed a strong agreement on every task shown in Figure 1, except for the fifth task, “validation of conversion rules.” The classification “transformation” was agreed upon for conformity. Based on the results, the agreed definition for the 3 terms was created in a consensus of all 10 reviewers.

Figure 1. Reviewers' categorization of the tasks of a metadata-driven data integration process. Red: matching; yellow: mapping; and blue: transformation.



Matching

The matching process describes the alignment of given data structures or metadata and creates an alignment proposal between the individual data elements. These matching candidates can be created by domain experts or matching algorithms by using equivalence classes (eg, equivalent, narrower, broader).

Mapping

In the mapping process, a domain expert uses the proposals of the matching process to define functions or uses external rule sets (eg, Unified Code for Units of Measure) to transform the source data structure into a target data structure. The conversion functions are not necessarily symmetrical.

Transformation

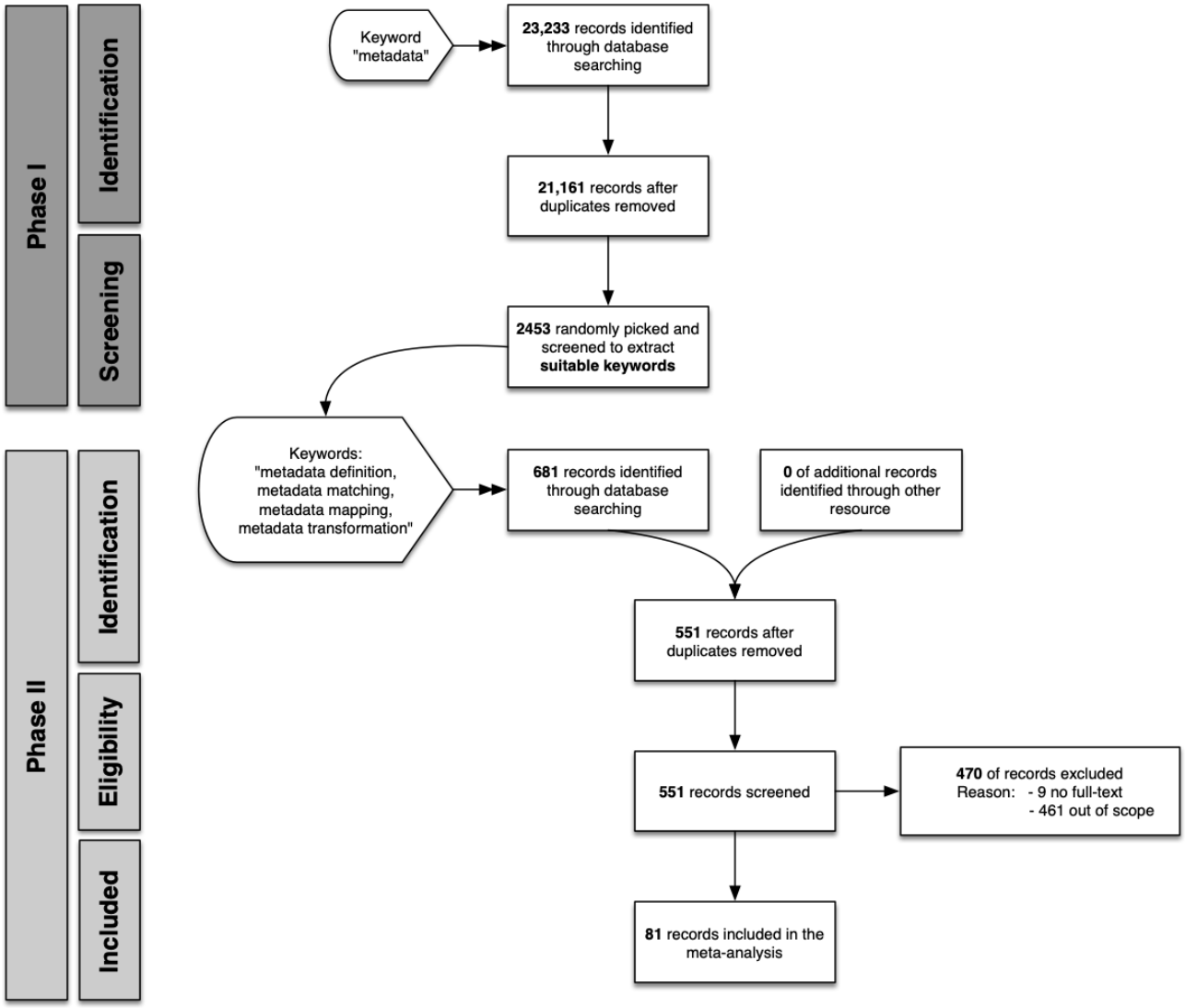
The transformation process combines metadata and instance data. It uses the conversion rules defined in the mapping process to transform instance data according to the target data structure.

Systematic Review

The first inquiry with the general keyword “metadata” was performed in mid-December 2019 and resulted in 23,233 papers—21,161 after duplication removal. Approximately 11.6% (2453/21,161) of the documents were randomly selected, resulting in 2453 publications whose titles and abstracts were

analyzed by the first author. The keywords of the relevant papers extended the search phrase to metadata definition, metadata matching, metadata mapping, and metadata transformation. The literature search was repeated in February 2020 using the extended search phrase in the second phase, resulting in 681 papers and 551 papers after removing the duplicated entries. The titles and abstracts were analyzed to match the scope by the first author, and 81 papers were selected for the full-text analysis (Figure 2). The papers were distributed across different disciplines: medical informatics (41papers), bibliography (10 papers), bioinformatics (8 papers), informatics (8 papers), social science (8 papers), geography (4 papers), neuroinformatics (1 paper), energy informatics (1 paper), and chemistry (1 paper). The review process was open for 8 weeks. The completed PRISMA checklist can be found as Multimedia Appendix 3. The results were gathered and analyzed by the first author and then discussed and approved by the reviewers.

Figure 2. The process for literature selection in 2 search phases with different keyword sets. Two separate literature inquiries were performed: the first inquiry aimed at identifying suitable keywords for the second literature inquiry, which provided papers for the full-text analysis.



Definition and Classification of Metadata

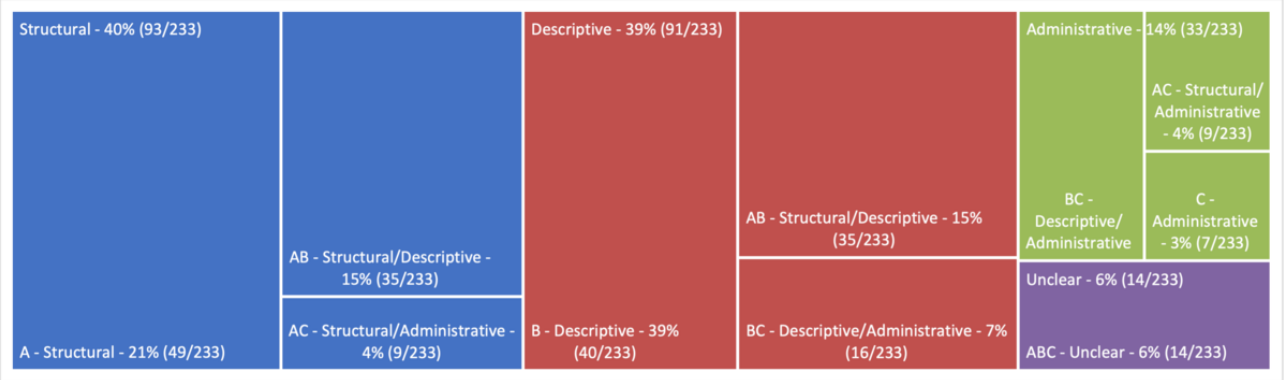
Guerra et al [12] stated that "metadata is an overloaded term in computer science and can be interpreted differently according to the context." The literature review confirmed this ideology, and the selected publications offered a variety of definitions. However, the general notion was that metadata is a formal representation of data that defines and describes information in a (preferable) standardized and stable way [13,14]. Various characteristics of this metadata definition were extracted from the publications:

- Small atomic units describing and constraining a specific object (table fields, attributes of form questions, records) [15]
- Describes data type, range, or set of possible values [16,17]
- Single units can be composed into complex elements [18]
- Single units are often called Data Element following the International Organization for Standardization (ISO) 11179 [19]
- Metadata can have bindings to terminologies, controlled vocabularies, and taxonomies [20,21]
- Metadata repositories or data dictionaries are used synonymously and store metadata centrally [16,22-24]

- Separation of content information from layout information [17]
- Detailed machine-readable and actionable descriptions to enable data processing without human guidance [10,25]

The NISO classification task showed that the majority of the papers were classified as structural or descriptive—papers with a pure focus on administrative metadata were a minority in the selected publications, as shown in Figure 3. The categorization of metadata according to the NISO has been described extensively elsewhere [26-29], but different definition schemes have also been encountered. Chu et al [30] introduced the separation of metadata with and without dependencies on the context. An important discriminator here is that some metadata capture information that is not dependent on the data. Context-independent metadata could describe more technical, provenance-specific records, whereas context-aware metadata could define the records descriptively to improve identification. The study from Grewe et al [31] described a new concept to annotate neurophysiological reports to capture as many annotations as possible. Therefore, the authors differed between *hard* and *soft* metadata. Parameters and information that could be directly measured (eg, temperature or timestamps) and assessed were called hard, whereas the reason of the experiment, the context information, and experiment rationale were labeled as soft metadata. Li et al [32] designed a data management system for a maritime observatory network and distinguished between 4 different metadata types: *data quality information* to ensure data reliability, *reference system information* to capture temporal and regional reference data, *maintenance information* to display updates and lifecycles, and *identification data*, which was the only mandatory type. A different categorization approach was chosen by Zozus and Bonner [33], which selected the described entity: record-level or data value–level metadata. This approach is particularly interesting for clinical studies, as the description at the value level, that is, the individual question fields in a study, is more conclusive than just a general description of the study.

Figure 3. The distribution of the publications included in this review. The categories were letter-encoded: A is structural, B is descriptive, and C is administrative, as well as their resulting combination. Structural (40%) and descriptive (39%) papers were clearly in the majority, while administrative (14%) papers were rarely found. Lastly, 6% of the papers could not be clearly classified.



Definitions of Matching, Mapping, and Transformation

Besides descriptions of metadata representations, some authors stated their understanding of metadata matching and mapping. Ashish et al [16] defined mapping as a one-to-one relationship across 2 data elements and a set of matching candidates as a suggestion window. Rebaï et al [34] described that mapping is a semantic correspondence relation between 2 metadata schemes, which have been identified in a schema matching process. Mate et al [35] shared this definition and considers mapping candidates as the result of a matching process. If a human expert approved the relation, a mapping candidate would become a mapping. In the study from Bernstein et al [36], a new differentiation was introduced: explicit and inferred mappings. An explicit or rather direct mapping was created between 2 metadata elements, whereas an inferred mapping used explicit mapping to create new relations like a metadata crosswalk [37]. Definitions of transformation were not found in the reviewed papers except in papers in which the reviewer coauthored [35,38]. The Fleiss kappa was calculated [39] for classification on the processing task to evaluate the interrater reliability, as seen in Table 1.

Table 1. The Fleiss kappa values to evaluate the interrater reliability of the classification task. Values between 0.00-0.20 are classified as slight agreement and values between 0.21-0.40 as fair agreement [39].

Task	Metadata processing task		
	Matching	Mapping	Transformation
Fleiss kappa	0.13175743	0.22358548	0.29233227

Used Metadata Standards

This review served to obtain insights into the standards and core data sets used. The assessments resulted in 37 relevant standards mentioned and used in the selected publications. The identified standards were grouped afterward into 3 categories following the levels of interoperability [40] for better oversight:

- Structure standards: ISO 11179, ISO 15926, ISO 19101, ISO 19763, ISO 20943, ISO 21526, ISO 23081, openEHR, CDISC ODM, OMOP, IHE DEX, Dublin Core, ASTM

CCR, CaDSR, EAD, GILS, VRA, CIMI, CSDGM, ONIX, MARC, TMA DES, EXIF, INSPIRE, SKOS, DCAT, W3C PROV
- Technical standards: XML, RDF, OWL, JSON-LD, ClaML
- Semantic standards: ICD-10, UMLS, SNOMED CT, LOINC, MedDRA, RxNorm.

Use Cases

Metadata are used for various use cases. The papers included in this review showed that metadata were mainly used for 4 tasks: information retrieval (21 papers), data integration (19 papers), core data set definition (10 papers), and the secondary use of data (7 papers). For information retrieval, metadata, especially semantic annotations, were used to improve query-based machine processing. Owing to a broader range of information descriptions, queries can be more accurately matched and thus, return more optional results. The processes of data integration and core data set definition used metadata to describe and harmonize the underlying schema, which can be used for secondary use of (eg, clinical) data. Further encountered use cases were an automatic data quality check [25,41] or ontology generation [42].

Problems and the Proposed Solutions

The reviewed papers addressed several problems regarding the processing and the use of metadata in different research fields and introduced solutions with new approaches to overcome obstacles. On analyzing the papers upon with described issues, we identified 5 problem categories: (1) structural-related problems, (2) semantics-related problems, (3) human interaction–related problems, (4) metadata lifecycle–related problems, and (5) metadata processing–related problems.

Structural-Related Problems

According to our review, the largest group of problems were structural-related issues. The authors of the reviewed papers described a lack of standard usage. They criticized a limited or confusingly extensive selection of suitable standards [41,42]. This affected the complexity of metadata [21] and data quality [36], which led to the underutilization of metadata [36]. The absence of standards and thus, their nonuse created several problems: metadata were heterogeneous in structure and format and contained bad or missing descriptions, preventing the understanding of existing metadata and resulting in low quality [43,44]. Using different units or precision for quantitative measurements complicated the usage [27], and the heterogeneous formats prevented machine readability, which therefore worsened the identification [45,46], accessibility [47], retrieval [31], and validation [26]. However, it must be emphasized that even the constant usage of standards did not avoid heterogeneity. Current standards have no extensibility functions to be future-proof [30] nor provide modularity to compose metadata blocks from different standards [48]. The commonly used standard ISO 11179 was no exception concerning those problems: missing hierarchical or temporal dependencies [13] and missing structural [49,50] or semantic extensions [51,52]. Several improvements concerning structural issues were found in the review: reducing ISO 11179 entities to streamline and improve ease-to-use [49], reconstructing the base models [53], or establishing a supermodel integrating all proprietary extensions and adaptions of the ISO 11179 [50]. A vast selection of standards was not conducive and foments metadata heterogeneity [22]. A good example is the field of bibliography, which has too many competing standards [54]. A possible way out of this standard jungle would be to reduce their amount by only using standards accepted by the research community [17] or reusing existing and validated data elements and definitions [25,55]. If no standards were suitable or the current method for defining standards was no longer appropriate, a new conceptual approach may help. Instead of creating new standards, Woodley [37] encouraged more investment in more effort in model agreement and model reconciliation. Corradi et al [18] described the use of an event-driven model to tackle the missing extensibility. Grewe et al [31] proposed a generic metamodel approach based on 5 characteristics: extensibility, modularity, refinements, multilingualism, and machine processability.

Semantics-Related Problems

Semantics is a big enabler for (meta)data reuse, and therefore, according to the literature, the lack of semantics was a difficult obstacle to overcome. A general problem related to every standardized data capture was the free-text elements [56]. Metadata elements also contained descriptions and definitions to understand the purpose of the items, but these included synonyms and spelling variations or naming conflicts [44], causing a data discrepancy problem if such data were shared. A viable solution was adding semantic codes to the corresponding data elements, which represented a deeper semantic understanding. Eichenlaub et al [44] assumed that de facto standard thesauri from research fields—in the authors' case fashion—did not cover (commonly) used terms, or the use of proprietary codes cause semantic heterogeneity [17]. The reviewed papers proposed a better annotation process, which a domain expert or natural language processing tools [23] should execute, supplemented by postcoordination and an expert review to ensure consistent encoding [57]. An essential addition would be the access and reuse of approved semantic annotations [20,58] or mapping property codes to standardized vocabularies [56].

The reviewed literature described another possible solution: the use of ontologies [15,59]. However, a problem with this approach was that an ontology must be created [60] or automatically constructed to match the instance data [61]. The reuse of existing ontologies and adaption to the custom requirements was likely a better and more adaptable choice [15]. However, problems arose when reusing ontologies owing to the metadata's necessary conformity with the ontology structure [62].

Human Interaction–Related Problems

The collaboration was described as an essential aspect mentioned in the reviewed papers from each research field. Sharing and discussing the created information was not only an opportunity to improve the designed data but a necessary step to overcome the hurdles of misinterpretation [48]. Human involvement was time- and resource-consuming owing to unfamiliar or complicated software, which resulted in a low

level of user acceptance [17]. Thus, metadata models or the corresponding software [63] were too complicated for health care professionals without certain necessary information technology skills [32] and therefore rarely used. In addition to the technical issues, the problem extended to the conceptual level: the model would not be clearly comprehensible if the stakeholder, users, and organizations slightly deviated in their understanding of the use cases [44]. As Varghese et al [55] aptly noted, simple disagreements about modeling decisions led to inadequate models. A tight feedback loop was recommended between users and the metadata curator to match the expected outcomes and a shared understanding of the metadata elements [44,64]. For example, extending metadata vocabularies with natural definitions would help to support the end users [64]. Nevertheless, vocabularies should be created with simplicity in mind and sufficiency instead of exhaustive description [65] as well as tooling. In the reviewed papers, 2 solutions were proposed. One approach stated that improved tools would enable medical experts for data modeling and a direct quality validation [17]. The second approach was to divide the work: the domain experts could deliver the knowledge, and metadata professionals would compose metadata in consultation, resulting in excellent and reusable metadata [66].

Metadata Lifecycle–Related Problems

Another vital issue is the divergence of data and the corresponding metadata [14]: data did not match the metadata and thus was not fit for reuse. The reasons for this were diverse: the lack of transparency of the (meta)data origin [47] or the boundary between data and metadata was unclear or rather a matter of changed perspective [28]. A viable approach was the extraction of metadata from the primary information technology systems and to populate it directly [23]. However, distributed metadata could vary across multiple data sources [67], and duplicates yielded the risk of staleness, particularly if the information was out-of-sync due to the extensive costs of metadata maintenance [51]. The reviewed publications state various measures that could be used against metadata staleness: continuous adaption and curation of metadata [43], tracking of changes during the metadata creation process [68], maintaining linkage information about provenance [69], and establishing a metadata lifecycle model [54]. Vos et al [70] pointed out a decisive circumstance: there is no current standard for archiving and preservation to cover the entire metadata lifecycle. However, especially archiving metadata was also the key to the reuse of archived data. Without the corresponding and descriptive metadata, the data would be difficult to reuse. Shean and Greninger [71] described that clinical metadata could even raise data privacy problems. Metadata may be used to infer other privacy-sensitive information. For example, metadata describing the parameter set specific for a HIV test connected to a particular patient could reveal the suspected disease and the diagnostic procedure to clarify the circumstances. Therefore, metadata should be considered to be anonymized before sharing to avoid data privacy concerns.

Metadata Processing–Related Problems

Metadata are often used for data harmonization to reduce labor. However, the process of metadata harmonization was usually performed manually [16], which was incredibly time- and resource-intensive [23]. Fortunately, the information was often machine-actionable, and therefore, automatic processing, especially matching and mapping, was possible. However, our literature review revealed known hurdles even before the metadata could be processed: heterogeneous metadata interfaces caused a siloization [72], which resulted in the impediment of metadata acquisition and reuse. If the information could be accessed, the processing also had problems: automatic matching from a broader to a more detailed level was nearly impossible [20], and if the matching results were promising, an automated mapping without human interaction was complicated or rather infeasible [16,24]. A stark problem resided in the fact that to improve the algorithms, more data for testing would be necessary, which were often challenging to obtain [38]. Moreover, the final merging of the data sets was also problematic: mappings could be ambiguous [29], the corresponding elements differed in the obligation level [73], or the proposed mapping had flaws and therefore, could cause information misinterpretation [53]. The reviewed papers proposed focusing on improving schema matching to enable a broader understanding of schemes [26]. The use of lexical and statistical methods would be enough for the matching process, and thus, the manual mapping afterward [38,57] would be indispensable to achieve adequate results. The matching could be refined with the use of unsupervised text mining techniques to calculate similarities between data elements [16]. To overcome the siloization of metadata, the use of standardized metadata search interfaces should be promoted and advanced, as shown by expanding Open Archives Initiative Protocol for Metadata Harvesting [74].

Discussion

Principal Results

The aim of this study was to investigate the anatomy of metadata and point out possible issues by conducting a deep insight into the recent academic literature in the last decade. It would have been desirable to extend the period to the previous 20 or even 30 years, but the amount of work would not be justifiable. The initial search for the actual review was intentionally broad with the generic key phrase “metadata,” resulting in 21,161 papers using Scopus and Web of Science. To maintain the general selection focus and minimize a self-imposed bias, domain-specific search engines such as PubMed were not used. Our selection criteria aimed for recent metadata papers with an emphasis on describing existing data sets to integrate them meaningfully. Papers dealing exclusively only with (instance) data or semantic standards were not included to reduce the immense amount of publications for review and concentrate on our core research interest. After several filtering steps, the resulting 81 papers included in the review were mainly from the field of medical informatics. This might be because metadata were very relevant to this area of research, and thus, a considerable amount of work was done in this area.

The papers’ distribution of the metadata categories was unbalanced: there were hardly any papers with an administrative orientation in the selected papers. The challenges of

comprehensible data collection and traceability intensified with a substantial increase in digitization, and administrative metadata can be used to support management processes. Intriguingly, this was apparently not strongly represented in the literature. This was somewhat surprising since this information would be indispensable for the documentation of origin and traceability of data records. It appeared that the field of administrative metadata, including provenance information, has been massively underrepresented in the last decade. The use cases found were in line with our daily experiences: metadata were mainly used to improve information retrieval and data integration. Another expected facet was the sheer amount of standards (see the comparative analysis of Baek and Sugimoto [54]). The multitude of different standards leads to oversaturation and rejection, which was an essential insight for medical informatics. Consequently, awareness of a limited number of supported standards that are improved and therefore followed by the community will be an important goal.

Besides the categorization of the NISO schema, other approaches were encountered. Upon closer inspection, the newly introduced models had a considerable overlap with the schema, except for 2 approaches. Chu et al [30] emphasize the focus on the context, which was not addressed within the NISO schema. The second approach was presented by Zozus and Bonner [33], which differentiated the described information by the level of detail. As the authors stated, especially in clinical trials, the fine granular definition of the data value level would be desirable. In contrast to the bibliography, where the entire record was essential for retrieval, in clinical studies, the question level was significant and should be defined and constrained as detailed as possible.

To ensure consistency, a harmonization process preceded our review. It had to be assured that all participating reviewers had the same understanding of the definitions. This harmonization step required additional time and effort but resulted in a joint set of definitions that could be evaluated during the review. To evaluate the differences in reviewers' understandings of these definitions, the Fleiss kappa was calculated. The results showed that the reviewers agreed on when metadata are used for mapping and transformation, although the process of matching had less agreement between experts. This can be explained by the partial mixing of the 2 definitions of matching and mapping in the analyzed publications, resulting in mixed results by the individual reviewer. The definitions and the differentiation between matching and mapping were congruent with the literature.

On the contrary, our understanding of transformation was divergent from the analyzed papers. Our definition was focusing on metadata-driven data integration: the usage of metadata for the transformation of (clinical) instance, whereas the found term *transformation* appears to be in the context of transforming the metadata itself. From this, an insight can be drawn: as a reviewer, we were influenced by our perspective on the context of metadata, and there was no consistent differentiation between *metadata transformation* and *instance data transformation*. Therefore, our definition could be used as a delimitation to define the latter field precisely.

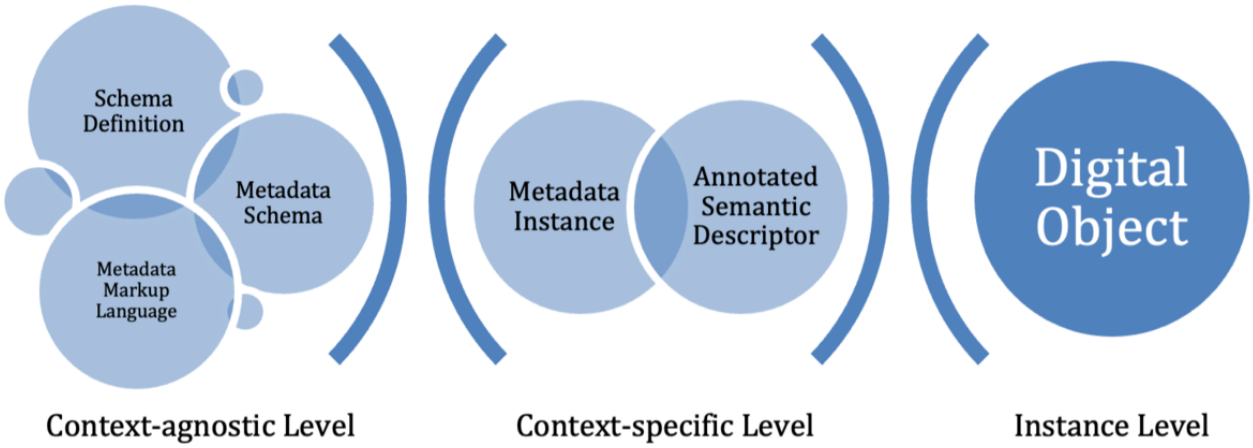
A further important insight was the dependence on context and perspective during the definition and evaluation of metadata, as Chu et al [30] designed their new model focusing on this fact. Consistent metadata require a high level of abstraction during its creation to be generally understood by the users. This would prevent inconsistent and incorrect (re)use of the metadata and the corresponding data. A related problem is known from the field of terminology engineering using different coding systems for the postcoordination [75]. The context influences the creation of information and blurs the precise line between structure and semantics. The information that should be universally applicable in the first place is affected by an individual point of view.

A New Schema Architecture of Metadata

Taking the decisive role of the metadata context into account, we derived a new schema for the classification of components for rich metadata objects adapting the model of Haslhofer and Klas [67]. As shown in Figure 4, the new scheme is based on the identification and separation of the context in terms of metadata creation. *Schema definition*, *metadata schema*, and a *metadata markup language* are context-agnostic. Representatives of each form the technical and semantic context in which a metadata object is instantiated. Metadata objects describe (non)digital objects. An illustration can be seen in Figure 3. Concerning context, the *metadata instance* can be enhanced using *annotated semantic descriptors* utilizing a variety of ontologies, terminologies, and coding systems. The instantiated metadata object can itself take the place of a digital object and be described in more detail by further metadata. This chaining mechanism allows a precise description of the highly networked nature of metadata. Further, chaining allows metadata from different systems and standards to be represented collectively in a single chained schema. An example is the enrichment of instance data with provenance information describing the origin of the metadata.

The schema definition can specify how metadata models are constructed. Well-known representatives are the norms ISO 11179 [19], ISO 15926 [76], and ISO 19763 [77]. The metadata schema describes the metadata objects with every needed attribute and is mostly the result of metadata harmonization and core data set creation, for example, Dublin Core or CaDSR. Metadata markup languages such as XML, RDF, or OWL are used for the technical description of the defined schema. The metadata schema and the metadata markup language are essential for metadata instantiation; the superimposed schema definition is not obligatory but highly recommended for comparability and interoperability.

Figure 4. The building blocks of metadata: schema definition, metadata schema, and markup language are jointly used to instantiate metadata with an additional semantic descriptor to describe a real-world object.



Limitations

This review showed that the term metadata *representation* is used as a synonym to the word *definition*, which could impact the analyzed paper selection. Furthermore, the initial paper selection could be a biased selection since the first author has a medical informatics background and was looking for a certain scope known from this. In addition, domain-specific search engines (such as PubMed) were not used; yet, the majority of papers were from the field of medical informatics. To avoid this bias, the initial selection could have been performed by various reviewers, but the sheer amount of work made this infeasible. It must also be mentioned that 10 papers were reviewed by only 2 persons because 1 reviewer had time constraints.

Comparison With Prior Work

To our knowledge, there is no comparable systematic review of metadata processing, which includes the analysis of approved solutions from other research fields and applicability to the field of medical informatics. Nevertheless, reviews on metadata have been carried out. Baek and Sugimoto [54] produced a review, which was included in our study, on existing metadata standards used in the bibliography community to identify the most suitable standard for electronic records. This review was limited to bibliography standards but gives an impressive overview. Singh and Bawa [78] analyzed techniques for metadata management and distribution in a large-scale storage system. This review focused only on the technical or administrative aspects of metadata. The newly introduced building block schema was adapted from Haslhofer and Klas [67], and additionally, the work from Ngouongo et al [53] must be mentioned. The study classified existing metadata formats to give a comparative overview and identify the most suitable candidate for the health care sector.

Conclusions

Metadata can be a powerful means to identify, describe, and process information, although its meaningful definition is challenging and entails significant hurdles. Different understanding of the same metadata representations is troublesome and hinders the correct utilization of metadata as well as the corresponding data instance. Through this work, 10 experts have gone through a consultation phase that ended in harmonized definitions for metadata in terms of metadata-driven data integration. This review process discovered many standards, use cases, problems, and solutions in dealing with metadata, providing a broad overview of the topic. This summary has led us to introduce a new schema for the classification of components for enriched metadata objects, which explicitly focuses on the creation context of metadata. These harmonized definitions and the new schema will improve the classification and creation of metadata by providing a mutual understanding of the metadata and its context.

Acknowledgments

We acknowledge financial support by Land Schleswig-Holstein within the funding program Open Access Publikationsfonds. Hannes Ulrich was funded by the German Research Foundation (Deutsche Forschungs-gemeinschaft) DFG grants IN 50/3-2. Jürgen Stausberg was funded by the German Federal Ministry of Education and Research under contract 01GY1917B. Martin Dugas was funded by the German Research Foundation grant DU 352/11-2.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The first survey form used for the harmonization process before the review.

[PDF File (Adobe PDF File), 220 KB - jmir_v24i1e25440_app1.pdf]

Multimedia Appendix 2
The second survey form used for the actual review process.
[PDF File (Adobe PDF File), 154 KB - jmir_v24i1e25440_app2.pdf]

Multimedia Appendix 3
The completed PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist for the review.
[DOCX File , 28 KB - jmir_v24i1e25440_app3.docx]

References

1. Ammenwerth E, Spötl HP. The time needed for clinical documentation versus direct patient care. A work-sampling analysis of physicians' activities. Methods Inf Med 2009;48(1):84-91. [Medline: 19151888]
2. Patel J. Bridging Data Silos Using Big Data Integration. IJDMS 2019 Jun 30;11(3):01-06 [FREE Full text] [doi: 10.5121/ijdms.2019.11301]
3. Hull R. Managing semantic heterogeneity in databases: a theoretical prospective. 1997 Presented at: Proceedings of the 16th ACM SIGACT-SIGMOD-SIGART symposium on Principles of database systems; May; NY, USA p. 51-61. [doi: https://doi.org/10.1145/263661.263668]
4. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. J Am Med Inform Assoc 2013 Jan 01;20(1):144-151 [FREE Full text] [doi: 10.1136/amiajnl-2011-000681] [Medline: 22733976]
5. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data 2016 Mar 15;3:160018 [FREE Full text] [doi: 10.1038/sdata.2016.18] [Medline: 26978244]
6. Vernadat FB. Technical, semantic and organizational issues of enterprise interoperability and networking. Annual Reviews in Control 2010 Apr;34(1):139-144. [doi: 10.1016/j.arcontrol.2010.02.009]
7. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. PLoS Med 2009 Jul 21;6(7):e1000100 [FREE Full text] [doi: 10.1371/journal.pmed.1000100] [Medline: 19621070]
8. Canakoglu A, Bernasconi A, Colombo A, Masseroli M, Ceri S. GenoSurf: metadata driven semantic search system for integrated genomic datasets. Database (Oxford) 2019 Jan 01;2019:baz132 [FREE Full text] [doi: 10.1093/database/baz132] [Medline: 31820804]
9. Gonçalves RS, Musen MA. The variable quality of metadata about biological samples used in biomedical experiments. Sci Data 2019 Feb 19;6:190021 [FREE Full text] [doi: 10.1038/sdata.2019.21] [Medline: 30778255]
10. Martínez-Romero M, O'Connor MJ, Egyedi A, Willrett D, Hardi J, Graybeal J, et al. Using association rule mining and ontologies to generate metadata recommendations from multiple biomedical databases. Database (Oxford) 2019 Jan 01;2019:baz059 [FREE Full text] [doi: 10.1093/database/baz059] [Medline: 31210270]
11. Understanding metadata: What is metadata, and what is it for? National Information Standards Organization. 2017. URL: https://www.niso.org/publications/understanding-metadata-2017 [accessed 2020-08-06]
12. Guerra E, Fernandes C. A Qualitative and Quantitative Analysis on Metadata-Based Frameworks Usage. 2013 Presented at: 13th International Conference in Computational Science and Its Applications – ICCSA 2013; June 24-27; Ho Chi Minh City, Vietnam. [doi: 10.1007/978-3-642-39643-4_28]
13. Löpprich M, Jones J, Meinecke M, Goldschmidt H, Knaup P. A reference data model of a metadata registry preserving semantics and representations of data elements. Stud Health Technol Inform 2014;205:368-372. [Medline: 25160208]
14. Li Z, Wen J, Zhang X, Wu C, Li Z, Liu L. ClinData Express--a metadata driven clinical research data management system for secondary use of clinical data. AMIA Annu Symp Proc 2012;2012:552-557 [FREE Full text] [Medline: 23304327]
15. Kim HH, Park YR, Lee KH, Song YS, Kim JH. Clinical MetaData ontology: a simple classification scheme for data elements of clinical data based on semantics. BMC Med Inform Decis Mak 2019 Aug 20;19(1):166 [FREE Full text] [doi: 10.1186/s12911-019-0877-x] [Medline: 31429750]
16. Ashish N, Dewan P, Toga AW. The GAAIN Entity Mapper: An Active-Learning System for Medical Data Mapping. Front Neuroinform 2015;9:30 [FREE Full text] [doi: 10.3389/fninf.2015.00030] [Medline: 26793094]
17. Breil B, Kenneweg J, Fritz F, Bruland P, Doods D, Trinczek B, et al. Multilingual Medical Data Models in ODM Format. Appl Clin Inform 2017 Dec 16;03(03):276-289. [doi: 10.4338/aci-2012-03-ra-0011]
18. Corradi L, Porro I, Schenone A, Momeni P, Ferrari R, Nobili F, et al. A repository based on a dynamically extensible data model supporting multidisciplinary research in neuroscience. BMC Med Inform Decis Mak 2012 Oct 08;12:115 [FREE Full text] [doi: 10.1186/1472-6947-12-115] [Medline: 23043673]
19. ISO/IEC 11179-7:2013 information technology-metadata registries (MDR). ISO. URL: https://www.iso.org/standard/50340.html [accessed 2020-02-18]

20. Daniel C, Sinaci A, Ouagne D, Sadou E, Declerck G, Kalra D. Standard-based EHR-enabled applications for clinical research and patient safety: CDISC-IHE QRPH-EHR4CR & SALUS collaboration. 2014 Presented at: AMIA Joint Summits on Translational Science; April 7-11; San Francisco p. 19.
21. Gonzalez-Beltran A, Campbell J, Dunn P, Guijarro D, Ionescu S, Kim H, et al. Data discovery with DATS: exemplar adoptions and lessons learned. J Am Med Inform Assoc 2018 Jan 01;25(1):13-16 [FREE Full text] [doi: 10.1093/jamia/ocx119] [Medline: 29228196]
22. Cunningham SG, Carinci F, Brillante M, Leese GP, McAlpine RR, Azzopardi J, Metelko, et al. Core Standards of the EUBIROD Project. Defining a European Diabetes Data Dictionary for Clinical Audit and Healthcare Delivery. Methods Inf Med 2016;55(2):166-176. [doi: 10.3414/ME15-01-0016] [Medline: 26666452]
23. Bruland P, Doods J, Storck M, Dugas M. What information does Your EHR contain? Automatic generation of a clinical metadata warehouse (CMDW) to support identification data access within distributed clinical research networks. 2017 Presented at: MEDINFO 2017: Precision Healthcare through Informatics; August 21-25; Hangzhou, China p. 313.
24. Kock-Schoppenhauer A, Bruland P, Kadioglu D, Brammen D, Ulrich H, Kulbe K, et al. Scientific Challenge in eHealth: MAPPATHON, a Metadata Mapping Challenge. Stud Health Technol Inform 2019 Aug 21;264:1516-1517. [doi: 10.3233/SHTI190512] [Medline: 31438209]
25. Lyttleton O, Wright A, Treanor D, Lewis P. Using XML to encode TMA DES metadata. J Pathol Inform 2011;2:40 [FREE Full text] [doi: 10.4103/2153-3539.84233] [Medline: 21969921]
26. Charles V, Isaac A, Fernie K, Dallas C, Gavrilis D, Angelis S. Achieving interoperability between the CARARE schema for monuments and sites and the Europeana Data Model. 2013 Presented at: International Conference on Dublin Core and Metadata Applications; September 2-6; Lisbon, Portugal p. 115 URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84891278291&partnerID=40&md5=08fce65ae38141e3d497b743186a1e14
27. Nadkarni P, Marenco L. Data integration: An overview. In: Methods in Biomedical Informatics. Amsterdam: Elsevier; 2013.
28. Papež V, Mouček R. Applying an Archetype-Based Approach to Electroencephalography/Event-Related Potential Experiments in the EEGBase Resource. Front Neuroinform 2017;11:24 [FREE Full text] [doi: 10.3389/fninf.2017.00024] [Medline: 28428750]
29. Song T, Park H, Jin D. Development of health information search engine based on metadata and ontology. Healthc Inform Res 2014 Apr;20(2):88-98 [FREE Full text] [doi: 10.4258/hir.2014.20.2.88] [Medline: 24872907]
30. Chu Y, Kuo W, Cheng Y, Lee C, Shiau C, Tarng D, et al. A Survival Metadata Analysis Responsive Tool (SMART) for web-based analysis of patient survival and risk. Sci Rep 2018 Aug 27;8(1):12880 [FREE Full text] [doi: 10.1038/s41598-018-31290-z] [Medline: 30150756]
31. Grewe J, Wachtler T, Benda J. A Bottom-up Approach to Data Annotation in Neurophysiology. Front Neuroinform 2011;5:16 [FREE Full text] [doi: 10.3389/fninf.2011.00016] [Medline: 21941477]
32. Li X, Yan T, Gao F, Zhou L, Yu J, Guo Z. Design of data management system for seafloor observatory network. 2013 Presented at: International Conference on Service Science; May 29-31; Washington p. 147. [doi: 10.1109/icss.2013.50]
33. Zozus M, Bonner J. Towards data value-level metadata for clinical studies. 2017 Presented at: Information Technology and Communications in Health; February 16-19; Victoria, BC, Canada p. 23. [doi: 10.3233/978-1-61499-742-9-418]
34. Rebaï RZ, Mnif F, Zayani CA, Amous I. Adaptive Global Schema Generation from Heterogeneous Metadata Schemas. Procedia Computer Science 2015;60:197-205. [doi: 10.1016/j.procs.2015.08.119]
35. Mate S, Kampf M, Rödle W, Kraus S, Proynova R, Silander K, et al. Pan-European Data Harmonization for Biobanks in ADOPT BBMRI-ERIC. Appl Clin Inform 2019 Aug;10(4):679-692 [FREE Full text] [doi: 10.1055/s-0039-1695793] [Medline: 31509880]
36. Bernstein M, Doan A, Dewey C. MetaSRA: normalized human sample-specific metadata for the Sequence Read Archive. Bioinformatics 2017 Sep 15;33(18):2914-2923 [FREE Full text] [doi: 10.1093/bioinformatics/btx334] [Medline: 28535296]
37. Woodley M. Crosswalks, metadata harvesting, federated searching, metasearching: Using metadata to connect users and information. Getty Research Institute. 2008. URL: https://getty.edu/research/publications/electronic_publications/intrometadata/path.pdf [accessed 2021-12-15]
38. Deppenwiese N, Duhm-Harbeck P, Ingenerf J, Ulrich H. MDRCupid: A Configurable Metadata Matching Toolbox. Stud Health Technol Inform 2019 Aug 21;264:88-92. [doi: 10.3233/SHTI190189] [Medline: 31437891]
39. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
40. Costin A, Eastman C. Need for Interoperability to Enable Seamless Information Exchanges in Smart and Sustainable Urban Systems. J. Comput. Civ. Eng 2019 May;33(3):04019008. [doi: 10.1061/(asce)cp.1943-5487.0000824]
41. Park J, Tosaka Y. Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms. Cataloging & Classification Quarterly 2010 Sep 27;48(8):696-715. [doi: 10.1080/01639374.2010.508711]
42. Huang G, Yuan M, Li C, Sun Q. Research on ontology generation and evaluation method in oil field based on the MDR. JCM 2017 Nov 24;17(4):665-676. [doi: 10.3233/jcm-170751]

43. De Jong F, Maegaard B, De Smedt K, Fišer D, Van Uytvanck D. Clarin: Towards fair and responsible data science using language resources. 2018 Presented at: International Conference on Language Resources and Evaluation; May; Miyazaki, Japan.
44. Eichenlaub N, Morgan M, Masak-Mida I. Undressing fashion metadata: Ryerson University Fashion Research Collection. 2014 Presented at: International Conference on Dublin Core and Metadata Applications; October 8-11; Austin URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85013637457&partnerID=40&md5=11880f4bdba4cec6af90520d38778437 [doi: 10.32920/ryerson.14637945.v1]
45. Ku HS, Kim S, Kim H, Chung H, Park YR, Kim JH. DialysisNet: Application for Integrating and Management Data Sources of Hemodialysis Information by Continuity of Care Record. Healthc Inform Res 2014 Apr;20(2):145-151 [FREE Full text] [doi: 10.4258/hir.2014.20.2.145] [Medline: 24872913]
46. Urban R. The 1:1 Principle in the age of Linked Data. 2014 Presented at: Proceedings of the International Conference on Dublin Core and Metadata Applications; October 8-11; Austin URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85013630440&partnerID=40&md5=6f2a602a3fe2d8c4fed4b4aab408d3bd
47. Maumet C, Auer T, Bowring A, Chen G, Das S, Flandin G, et al. Sharing brain mapping statistical results with the neuroimaging data model. Sci Data 2016 Dec 06;3:160102 [FREE Full text] [doi: 10.1038/sdata.2016.102] [Medline: 27922621]
48. Trani L, Atkinson M, Bailo D, Paciello R, Filgueira R. Establishing Core Concepts for Information-Powered Collaborations. Future Generation Computer Systems 2018 Dec;89:421-437. [doi: 10.1016/j.future.2018.07.005]
49. Milward D. Model driven data management in healthcare. 2019 Presented at: International Conference on Model-Driven Engineering and Software Development; February 20-22; Prague, Czech Republic p. 105-117. [doi: 10.5220/0007391101070118]
50. Park Y, Kim J. Achieving interoperability for metadata registries using comparative object modeling. Stud Health Technol Inform 2010;160(Pt 2):1136-1139. [Medline: 20841861]
51. Ngouongo S, Stausberg J. Integration of classifications and terminologies in Metadata registries based on ISO/IEC 11179. Stud Health Technol Inform 2011;169:744-748. [Medline: 21893846]
52. Park YR, Yoon YJ, Kim HH, Kim JH. Establishing semantic interoperability of biomedical metadata registries using extended semantic relationships. Stud Health Technol Inform 2013;192:618-621. [Medline: 23920630]
53. Ngouongo S, Löbe M, Stausberg J. The ISO/IEC 11179 norm for metadata registries: does it cover healthcare standards in empirical research? J Biomed Inform 2013 Apr;46(2):318-327 [FREE Full text] [doi: 10.1016/j.jbi.2012.11.008] [Medline: 23246614]
54. Baek J, Sugimoto S. A task-centric model for archival metadata schema mapping based on the records lifecycle. IJMSO 2012;7(4):269. [doi: 10.1504/ijmso.2012.051488]
55. Varghese J, Fujarski M, Hegselmann S, Neuhaus P, Dugas M. CDEGenerator: an online platform to learn from existing data models to build model registries. CLEP 2018 Aug;Volume 10:961-970. [doi: 10.2147/clep.s170075]
56. Berry S, Edgar H. Standardizing Data from the Dead. Stud Health Technol Inform 2019 Aug 21;264:1427-1428. [doi: 10.3233/SHTI190467] [Medline: 31438164]
57. Pathak J, Wang J, Kashyap S, Basford M, Li R, Masys DR, et al. Mapping clinical phenotype data elements to standardized metadata repositories and controlled terminologies: the eMERGE Network experience. J Am Med Inform Assoc 2011 Jul 01;18(4):376-386. [doi: 10.1136/amiajnl-2010-000061]
58. Dugas M, Hegselmann S, Riepenhausen S, Neuhaus P, Greulich L, Meidt A, et al. Compatible Data Models at Design Stage of Medical Information Systems: Leveraging Related Data Elements from the MDM Portal. Stud Health Technol Inform 2019 Aug 21;264:113-117. [doi: 10.3233/SHTI190194] [Medline: 31437896]
59. Hall S, McMahon B. The implementationevolution of STAR/CIF ontologies: Interoperability preservation of structured data. Data Science Journal 2016;15:3. [doi: 10.5334/dsj-2016-003]
60. Lunesu M, Pani F, Concas G. Using a standards-based approach for a multimedia knowledge-base. 2011 Presented at: International Conference on Knowledge Management and Information Sharing; October 26-29; Paris, France p. 87. [doi: 10.5220/0003663300870095]
61. Ivanschitz B, Lampoltshammer T, Mireles V, Revenko A, Schlarb S, Thurnay L. A semantic catalogue for the data market Austria. 2018 Presented at: 2nd Workshop on Decentralizing the Semantic Web co-located with the 17th International Semantic Web Conference; September 10-13; Monterey URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-85053674305&partnerID=40&md5=229a4cfc12b56c406c825457b7d58941
62. Frosini L, Bardi A, Manghi P, Pagano P. An aggregation framework for digital humanities infrastructures: The parthenos experience. SCIRES-IT 2018;8:50. [doi: 10.2423/i22394303v8n1p33]
63. Späth MB, Grimson J. Applying the archetype approach to the database of a biobank information management system. Int J Med Inform 2011 Mar;80(3):205-226. [doi: 10.1016/j.ijmedinf.2010.11.002] [Medline: 21131230]
64. Howarth LC. Designing a Common Namespace for Searching Metadata-Enabled Knowledge Repositories: An International Perspective. Cataloging & Classification Quarterly 2003 Jul;37(1-2):173-185. [doi: 10.1300/j104v37n01_12]
65. Rodrigues J, Castro J, da SJ, Ribeiro C. Hands-On Data Publishing with Researchers: Five Experiments with Metadata in Multiple Domains. Digital Libraries: Supporting Open Science 2019;988:88. [doi: 10.1007/978-3-030-11226-4_22]

66. Kock-Schoppenhauer A, Kroll B, Lambarki M, Ulrich H, Stahl-Toyota S, Habermann J, et al. One Step Away from Technology but One Step Towards Domain Experts-MDRBridge: A Template-Based ISO 11179-Compliant Metadata Processing Pipeline. Methods Inf Med 2019 Dec;58(S 02):e72-e79 [FREE Full text] [doi: 10.1055/s-0039-3399579] [Medline: 31853911]
67. Haslhofer B, Klas W. A survey of techniques for achieving metadata interoperability. ACM Comput. Surv 2010 Feb;42(2):1-37. [doi: 10.1145/1667062.1667064]
68. Park J, Tosaka Y. Metadata Creation Practices in Digital Repositories and Collections: Schemata, Selection Criteria, and Interoperability. ITAL 2010 Sep 01;29(3):104. [doi: 10.6017/ital.v29i3.3136]
69. Francis W, Atkinson R, Box P, Rankine T, Woodman S, Kostanski L. Model-driven data harvesting to publish provenance for geospatial references. 2013 Presented at: International Conference on Knowledge Capture: Knowledge Capture in the Age of Massive Web Data; June 23-26; Banff, Canada. [doi: 10.1145/2479832.2479863]
70. Vos R, Balhoff J, Caravas J, Holder M, Lapp H, Maddison W, et al. NeXML: rich, extensible, and verifiable representation of comparative data and metadata. Syst Biol 2012 Jul;61(4):675-689 [FREE Full text] [doi: 10.1093/sysbio/sys025] [Medline: 22357728]
71. Shean R, Greninger A. Private collection: high correlation of sample collection and patient admission date in clinical microbiological testing complicates sharing of phylodynamic metadata. Virus Evol 2018 Jan;4(1):vey005 [FREE Full text] [doi: 10.1093/ve/vey005] [Medline: 29511571]
72. Jeong S, Kim HH, Park YR, Kim JH. Clinical Data Element Ontology for Unified Indexing and Retrieval of Data Elements across Multiple Metadata Registries. Healthc Inform Res 2014 Oct;20(4):295-303 [FREE Full text] [doi: 10.4258/hir.2014.20.4.295] [Medline: 25405066]
73. Specka X, Gärtner P, Hoffmann C, Svoboda N, Stecker M, Einspanier U, et al. The BonaRes metadata schema for geospatial soil-agricultural research data – Merging INSPIRE and DataCite metadata schemes. Computers & Geosciences 2019 Nov;132:33-41. [doi: 10.1016/j.cageo.2019.07.005]
74. Yuliant R, Karna N. Knowledge sharing filtering on OAI-PMH. 2016 Presented at: International Conference on Information Technology Systems and Innovation, ICITSI; October 24-27; Bandung, Bali, Indonesia. [doi: 10.1109/icitsi.2016.7858213]
75. Qamar R, Kola J, Rector A. Unambiguous data modeling to ensure higher accuracy term binding to clinical terminologies. 2007 Presented at: AMIA Annual Symposium; November 10-14; Chicago.
76. ISO 15926-13:2018 Industrial automation systems and integration — Integration of life-cycle data for process plants including oil and gas production facilities — Part 13: Integrated asset planning life-cycle. ISO. URL: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/07/06/70694.html [accessed 2020-08-18]
77. ISO/IEC 19763-1:2015 Information technology — Metamodel framework for interoperability (MFI) — Part 1: Framework. ISO. URL: https://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/46/64637.html [accessed 2020-02-18]
78. Singh HJ, Bawa S. Scalable Metadata Management Techniques for Ultra-Large Distributed Storage Systems -- A Systematic Review. ACM Comput. Surv 2018 Sep 06;51(4):1-37. [doi: 10.1145/3212686]

Abbreviations

ISO: International Organization for Standardization
NISO: National Information Standards Organization
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by R Kukafka, G Eysenbach; submitted 02.11.20; peer-reviewed by G Zhang, X Jing, T Sagi; comments to author 22.12.20; revised version received 28.01.21; accepted 14.10.21; published 11.01.22.

Please cite as:
Ulrich H, Kock-Schoppenhauer AK, Deppenwiese N, Gött R, Kern J, Lablans M, Majeed RW, Stöhr MR, Stausberg J, Varghese J, Dugas M, Ingenerf J
Understanding the Nature of Metadata: Systematic Review
J Med Internet Res 2022;24(1):e25440
URL: https://www.jmir.org/2022/1/e25440
doi:10.2196/25440
PMID:35014967

Review

Barriers to and Facilitators of Automated Patient Self-scheduling for Health Care Organizations: Scoping Review

Elizabeth W Woodcock[1], BA, MBA, DrPH

Department of Health Policy & Management, Rollins School of Public Health, Emory University, Atlanta, GA, United States

Corresponding Author:
Elizabeth W Woodcock, BA, MBA, DrPH
Department of Health Policy & Management
Rollins School of Public Health
Emory University
1518 Clifton Road
Atlanta, GA, 30307
United States
Phone: 1 404 272 2274
Email: elizabeth@elizabethwoodcock.com

Abstract

Background: Appointment management in the outpatient setting is important for health care organizations, as waits and delays lead to poor outcomes. Automated patient self-scheduling of outpatient appointments has demonstrable advantages in the form of patients' arrival rates, labor savings, patient satisfaction, and more. Despite evidence of the potential benefits of self-scheduling, the organizational uptake of self-scheduling in health care has been limited.

Objective: The objective of this scoping review is to identify and to catalog existing evidence of the barriers to and facilitators of self-scheduling for health care organizations.

Methods: A scoping review was conducted by searching 4 databases (PubMed, CINAHL, Business Source Ultimate, and Scopus) and systematically reviewing peer-reviewed studies. The Consolidated Framework for Implementation Research was used to catalog the studies.

Results: In total, 30 full-text articles were included in this review. The results demonstrated that self-scheduling initiatives have increased over time, indicating the broadening appeal of self-scheduling. The body of literature regarding intervention characteristics is appreciable. Outer setting factors, including national policy, competition, and the response to patients' needs and technology access, have played an increasing role in influencing implementation over time. Self-scheduling, compared with using the telephone to schedule an appointment, was most often cited as a relative advantage. Scholarly pursuit lacked recommendations related to the framework's inner setting, characteristics of individuals, and processes as determinants of implementation. Future discoveries regarding these Consolidated Framework for Implementation Research domains may help detect, categorize, and appreciate organizational-level barriers to and facilitators of self-scheduling to advance knowledge regarding this solution.

Conclusions: This scoping review cataloged evidence of the existence, advantages, and intervention characteristics of patient self-scheduling. Automated self-scheduling may offer a solution to health care organizations striving to positively affect access. Gaps in knowledge regarding the uptake of self-scheduling by health care organizations were identified to inform future research.

(J Med Internet Res 2022;24(1):e28323) doi:10.2196/28323

KEYWORDS

appointment; scheduling; outpatient; ambulatory; online; self-serve; e-book; web-based; automation; patient satisfaction; self-scheduling; eHealth; digital health; mobile phone

Introduction

Background

Appointment management in the outpatient setting is important for health care organizations, as waits and delays lead to poor outcomes. The Institute of Medicine has 6 aims for health care organizations to improve quality [1]. Despite the goal of timely access to care, the topic of visit timeliness is one of the least evaluated and understood aspects of care delivery, and there is little assessment of what drives care timeliness and the potential approaches for improving this dimension of care [2].

Appointment wait times and scheduling difficulties can negatively affect patient satisfaction [3-5], access to care [6], patient safety [7], and health care use and organizational reputation [2]. Timely access has a broader impact on the delivery of cost-effective health care [8] and individuals' well-being [9]. The association between patient experience and the perception of quality of care has been demonstrated by Schneider et al [10]. Reasonable wait times are expected by patients [11,12].

Outside of health care, other industries with limited resources have addressed timeliness to service by engaging customers through self-service. For example, the transportation and hospitality industries have experienced improvements in operations [13,14], profitability [15], customer loyalty [16], and customer wait times [17] via the execution of consumer-based reservation systems. At present, consumers make reservations for services from a multitude of non–health care businesses. However, the adoption of management technologies, such as the self-scheduling of appointments in health care, has trailed other industries.

Benefits

There is evidence that automated self-scheduling provides value and that health care organizations can benefit from it. Researchers have identified the advantages of automated patient self-scheduling for health care organizations in the form of labor savings [18-22], information transparency [23,24], cost reduction [25], cycle time [26], patient satisfaction [27,28], patient accountability [29], patient information [30], patient time savings [31], physician punctuality [32], patient loyalty [23], and patient attendance [33-37]. Reducing missed appointments increases a health care organization's efficiency and the effective allocation of resources [38]. Automated self-scheduling eliminates the barriers inherent in the fixed capacity of phone lines and scheduling staff [39].

Health care organizations are faced with the need to increase access to accommodate patients' changing expectations [40,41]. Self-scheduling may offer the convenience that patients seek [42,43]. Countries in Europe [44], England [19,34], Canada [36], Australia [45], and the United States [23], have established health technology initiatives at the national level. Nigeria [20], India [30], Taiwan [22], the Philippines [26], and Iraq and the Kurdistan region [46] have determined that self-scheduling may serve as a better alternative to obtaining an appointment as opposed to the traditional process of accessing outpatient care by physically standing in line. In Iran [47] and China [24,48,49], hospitals are mandated to provide the capability, in part, to address the problems associated with in-person queues for appointments. In Estonia, this functionality is built into the national system [50]. The benefits of self-scheduling may not be realized by persons in low- and middle-income countries, where many patients report negative experiences related to poor communication, short visits, or lengthy waits [51]. Self-scheduling may be perceived as elusive or ineffective, with patients preferring to physically wait in line to combat inefficiencies. This may not be a malfunction of the technological solution but rather a result of low- and middle-income countries' failure to address socioeconomic disparities that have eroded patients' confidence in the health care system [52].

Adoption

Despite evidence of the potential benefits of self-scheduling, the organizational uptake of self-scheduling in health care has been limited. The lack of adoption may be a result of several factors examined in other studies of technology adoption, including the absence of financial incentives for the organization [53], cost [54], leadership [55], and policy and regulations [56]. Health care providers have expressed reluctance about self-scheduling based on cost, flexibility, safety, and integrity; patients cited concerns based on their prior experience with computers and the internet, as well as communication preferences [21]. Organizations may be reacting to patient hesitancy. Despite the infusion of technology in daily living, patients exhibit reluctance to automation in health care, citing concerns about accuracy, security, and the lack of empathy compared with human interactions [57].

There is a small body of literature regarding organizational barriers to the adoption of automated self-scheduling in popular literature. A practicing physician, informaticist, and the founder of a software company that offered self-scheduling products, Dr Jonathan Teich, revealed the following to the American Medical News in 2004 [58]:

> *Before you can successfully implement self-scheduling, you have to implement "Mabel." Mabel is the generic scheduling administrator who has been working for Dr. Smith for 35 years, and knows a thousand nuances and idiosyncrasies and preferences that have been silently established over the years...Unfortunately for the computer world, it's extremely difficult to find out what Mabel really knows, let alone try and put it into an algorithm.*

Research has demonstrated that physicians' concerns about addressing scheduling complexity [58] and preferences [59] are key factors in scheduling, with physicians expressing a fear of losing control of their schedules [60-62].

A previous review in this field provided evidence of facilitators of (no-shows, labor, waiting time, and patient satisfaction) and barriers to (cost, flexibility, safety, and integrity) automated self-scheduling [21]. Patients' expectations regarding their health care experience, as well as the application, adoption, and use of health care technology have evolved significantly since the publication of the systematic review in 2017, thereby compelling a new review to be performed.

Aim

Against this background, this scoping review seeks to identify the barriers to and facilitators of self-scheduling for health care organizations. The scoping review technique was selected based on a broad research question, the pursuit of identifying content without judging the quality of the material, and the intention to perform a qualitative synthesis [63].

Methods

The five-step process for scoping reviews by Arksey and O'Malley [64] was deployed for this study: (1) identification of the research question; (2) identification of relevant studies; (3) study selection; (4) charting the data; and (5) collating, summarizing, and reporting the results.

Step 1: Identification of the Research Question

The following research questions guided the review: What are the barriers to and facilitators of health care organizations' uptake of automated patient self-scheduling? What are the gaps in the literature regarding barriers and facilitators?

Step 2: Identification of Relevant Studies

This scoping review was performed by searching electronic databases according to the PRISMA-S (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Search) guidelines [65]. The databases used were PubMed, CINAHL, Business Source Ultimate, and Scopus. The search strategy was developed with the assistance of an informaticist specializing in reviews. The search terms for self-scheduling were developed by researching titles, keywords, and commonly used phrases in the relevant literature. The search strategy was initiated on PubMed using combinations and word variations of key terms for the scoping review: "self-scheduling," "automated scheduling," "Web-based scheduling," "e-appointments," "online scheduling," "Internet scheduling," and "self-serve scheduling." Additional terms were integrated using keywords from articles of interest that were retrieved from a preliminary search on PubMed. The implementation-related search string was adapted from a study of barriers and facilitators [66]. The initial search strategy was referenced against the published systematic review by Zhao et al [21] to identify supplementary terms. The search strategies used in the databases are reported in Multimedia Appendix 1. Articles were identified, screened, and selected for further review in two stages by the author: titles and abstracts, followed by the full text.

Step 3: Study Selection

Records were selected if they involved automated patient self-scheduling. Articles were determined eligible for inclusion if they discussed the use of self-scheduling by health care organizations. Peer-reviewed articles, primary research, reviews, and original studies described in editorials in peer-reviewed journals that focused on patient self-scheduling were included. Only articles published in English were included during study selection.

For the review, the definition of self-scheduling involves real time, synchronous booking, and automated fulfillment of appointments by patients on the web or via a smartphone app for themselves. Self-scheduling does not include an appointment by a physician on behalf of a patient, as in the case of a primary care physician scheduling an appointment with a specialist for the patient. Furthermore, the definition excludes asynchronous scheduling transactions that feature the patient initiating a request for an appointment but not booking it automatically, or the slot being appointed automatically through a waitlist feature [67] or a reschedule option [68]. Patients scheduled as research participants were excluded. The definition excludes self-scheduling of providers and staff.

Step 4: Charting the Data

A data extraction Microsoft Excel spreadsheet was developed to systematically record the details of the articles. Charted data (Multimedia Appendix 2 [5, 18, 19, 21-24, 26, 28, 29, 31-37, 42-45, 47, 48, 69-75]) included article characteristics (author, year, and country), intervention characteristics (stand-alone or component, source, introduction, description of design, and identified need), research design, setting, intervention measures assessing the impact of self-scheduling, and main results. Relevant results were extracted from the results section of each article.

Step 5: Collating, Summarizing, and Reporting the Results

The scoping review was organized and presented in alignment with the Consolidated Framework for Implementation Research (CFIR). The conceptual framework provides guidance for the research by constructing a standard, evidence-based path for identifying, organizing, and communicating the dimensions of barriers and facilitators across organizations to advance the opportunity for adoption of the study's findings. The framework is comprehensive, synthesizing essential constructs from 29 organizational and implementation science theories. Standard terminology promotes generalizability across disciplines (Textbox 1) [76].

Thematic analysis was performed to convey the main findings of the material.

Textbox 1. Consolidated Framework for Implementation Research domains and constructs.

Intervention characteristics

- Intervention source
- Evidence strength and quality
- Relative advantage
- Adaptability
- Trialability
- Complexity
- Design quality and packaging
- Cost

Outer setting

- Patient needs and resources
- Cosmopolitanism
- Peer pressure
- External policy and incentives

Inner setting

- Structural characteristics
- Networks and communications
- Culture
- Implementation climate
- Readiness for implementation

Characteristics of individuals

- Knowledge and beliefs about the intervention
- Self-efficacy
- Individual stage of change
- Individual identification with organization
- Other personal attributes

Process

- Planning
- Engaging
- Executing
- Reflecting and evaluating

Results

Overview

Titles and abstracts were reviewed for 1726 records, with 1604 (92.93%) records being excluded. The full texts of 7.06% (122/1726) of articles were retrieved and reviewed. In total, 5.33% (92/1726) of studies were excluded because they failed to meet the inclusion criteria. A total of 1.73% (30/1726) of studies were included in this scoping review. Figure 1 outlines the selection methodology using a PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) diagram.

The countries covered in the review include the United States [18,28,29,31,33,35,37,43,73,74], Taiwan [22,23,42], England [19,34,69,72], China [24,48,75], Australia [45,70,71], Canada [36], Iran [5,32,47], and the Philippines [26]. Another article included 7 countries in Europe [44]. Table 1 presents the countries and the number of articles from each. The first article retrieved for the scoping study was published in 2004 [18], with ≤3 articles each year up to and including 2019. In 2020, 8 articles [22,26,28,29,31,37,72,73] featuring barriers to and facilitators of automated self-scheduling were published. Table 2 displays the number of articles published by year of publication.

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

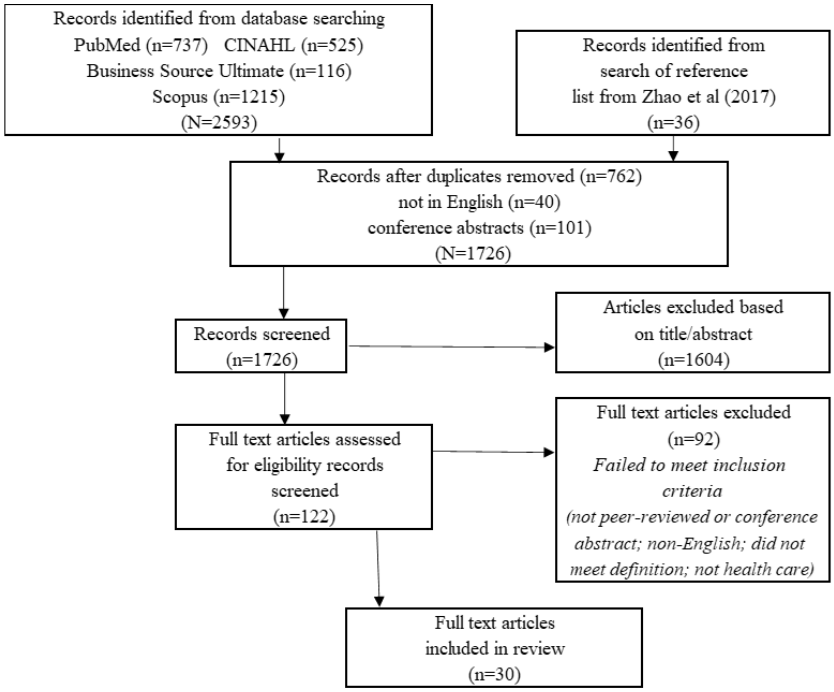


Table 1. Country-wise number of articles published (N=30).

Country	Articles, n (%)
United States	10 (33)
England	4 (13)
Taiwan	3 (10)
China	3 (10)
Australia	3 (10)
Iran	3 (10)
Canada	1 (3)
Philippines	1 (3)
7 countries in Europe	1 (3)
Other (review)	1 (3)

Table 2. Articles by year of publication (N=30).

Year	Articles, n (%)
2004	1 (3)
2005	0 (0)
2006	0 (0)
2007	1 (3)
2008	1 (3)
2009	1 (3)
2010	2 (7)
2011	2 (7)
2012	1 (3)
2013	2 (7)
2014	3 (10)
2015	1 (3)
2016	0 (0)
2017	2 (7)
2018	3 (10)
2019	2 (7)
2020	8 (26)

Intervention Characteristics

Intervention Source

Of the 30 articles selected, 4 (13%) articles reported internal solutions for self-scheduling [28,31,37,73]. In addition to these studies, 7% (2/30) of articles were included that were published with a combination of internal and external resources [18,19]. From the 30 articles, 6 (20%) articles featured externally created interventions, 4 (13%) of which were created by a third party [35,36,43,72], 1 (3%) by the first author [71], and 1 (3%) by an unknown source [34]. The remaining articles did not elucidate the source of the intervention [5,22-24,26,32,42,44,45,47,48,70], did not feature a specific source [29,33,69,74,75], or represented a systematic review [21].

In total, of the 30 articles, 9 (30%) [18,19,22,26,32,34,45,48,72] provided some level of description of the intervention, with 4 (13%) providing only limited characteristics [22,26,34,72]. Most articles [5,18,23,24,26,32-37,43,45,47,48,70,71,73] featured the self-scheduling intervention as a stand-alone service, with a minority [19,22,28,29,31,42,44,72,74-76] including self-scheduling as a component of a larger technology offering. A systematic review [21] discussed self-scheduling in both contexts. The literature includes limited information regarding the source of the intervention. Sources were not cited as a barrier to or facilitator of implementation. This is evidenced by the volume of unknown and undescribed sources. The internally developed solutions, all reported in 2020, may imply that there is easier access for health care organizations to implement self-scheduling solutions.

Evidence Strength and Quality

The measurement of outcomes was a prominent element of the articles; however, the strength and quality of evidence was not presented as a determinant in the implementation of self-scheduling by the organization. The systematic review concluded that researchers demonstrated a reduced no-show rate, decreased staff labor, decreased waiting time, and improved patient satisfaction [21]. In the literature, evidence has not been measured on a consistent basis. For example, a case study documented a specific reduction in costs: a decrease of 25% of staff dedicated to scheduling, with an annual savings of US $170,000 for the organization [18]. The specifics of the roles of those personnel, their compensation, or other factors were not reported. Another study [72] reported on the intervention's *anticipated* results. The literature did not provide a robust body of evidence that may have influenced the implementation of self-scheduling by health care organizations.

Relative Advantage

The advantages of the intervention compared with alternative solutions have been discussed in the literature. The comparison was made with the option of using a telephone to schedule an appointment [18,19,28,31,33,37,45,69,70,72-75]. The literature revealed the relative advantage of self-scheduling being the use of the solution at any hour to overcome patient barriers to scheduling appointments [69,72]. The findings reported that 34% [45], 46% [37], and 51% [19] of appointments were self-scheduled outside of office hours. After-hours access to the health care organization allowed early morning appointments to be filled, thus benefiting the organization [33]. In their findings, studies detailed an improved use of staff resources [18,28,37,45,70,73,75] and time savings for the patient

[19,31,74]. Volk et al [28] hypothesized that self-scheduling offered patients an enhanced sense of anonymity and a diminished sense of responsibility, compared with the traditional telephone-based scheduling process.

Adaptability and Trialability

Faced with a surge in patient demand owing to the COVID-19 pandemic, an organization rapidly introduced the intervention [31]. This implementation provided evidence of adaptability and trialability as determinants that promoted the implementation of self-scheduling. The importance of allowing each practice the latitude to adopt their own strategy for marketing the intervention was observed; in 1 health care organization, by the second year of adoption, 20% of all slots were booked via self-scheduling [36]. Without any promotion, researchers observed a 300% increase in self-scheduled appointments within months [19]. The rapidity of implementation, customization of the solution, and patient use without promotion provide evidence of the determinants of adaptability and trialability to facilitate implementation.

Complexity

Although most of the studies did not describe the intervention, several studies made note of elements that revealed the complexity of the intervention. Slot unavailability was cited as a deterrent for patients attempting to self-schedule [36,45]. Ease of use was confirmed to be a key attribute for self-scheduling from the perspective of the patient [23,29]. These findings contrast those of Lee et al [22], who concluded that ease of use was not a facilitating attribute; instead, the researchers ascertained that performance expectancy was the determinant. Solutions that were bundled with triage featured an algorithm that diverted patients with acute symptoms from the self-scheduling option [19,31]. In all, 3% (1/30) of organizations reviewed appointments manually for safety and appropriateness [73]. The complexity of the intervention was reported to be important to manage [33], suggesting that it is a determinant of implementation success for health care organizations.

Design Quality and Packaging

The literature did not elaborate on the design quality and packaging of the intervention, except for sample screenshots of the patient interface [18,32,69]. Studies have highlighted the importance of integration with other information technology systems [33,36]. In all, 3% (1/30) of studies pointed out a predetermined lack of publicity: a health care organization during the pandemic avoided promotion to prevent artificially inducing additional patient demand [31]. A key factor in adoption was the organization making patients aware of the intervention [36,45,70]. Brochures made available to patients were reported to be ineffective in raising awareness [36,70,71]. Health care organizations documented the importance of presenting self-scheduling to patients using communication methods planned locally, as varying methods of approach may affect outcomes [72-74].

Cost

Although concern about cost was revealed as a barrier to physicians' interest in offering self-scheduling, information about the cost of the intervention to the health care organization was not addressed in the literature [21]. One author funded the intervention personally [18]. However, no details were provided regarding the amount spent for the intervention.

Outer Setting

Patient Needs and Resources

Concerns have been raised regarding possible disparities in care access for Medicaid recipients in the United States owing to lower provider count and longer distance to appointments via third-party self-scheduling platforms [43], as well as lower use rates of self-scheduling compared with non-Medicaid patients [31,73]. Research has provided evidence of diminished access to self-scheduling for rural patients compared with urban patients [35]. Low socioeconomic status was a driver of low adoption rates [45,72], with younger [19,37,45,72,73] women [19,37] who were employed [45] and patients with higher education [24,45] using the self-scheduling platform. Younger patients expressed the value of self-scheduling, as compared with users more senior to them [44,48,74]. One study [34] concluded that older patients were higher users; their study focused on the self-scheduling of specialty visits following a primary care physician's referral, thereby indicating that patients were specifically instructed to self-schedule. Patients with comorbidities were shown to be more frequent users than other patients [73]. Although most studies measured patient awareness, characteristics, use, and intention to use, there has been a growing interest over time in accounting for patients' needs and resources.

Multiple studies identified patients' access to the internet and computers as a potential barrier to the use of self-scheduling [35,45,70,71,74]. In a postintervention focus group, Mendoza et al [26] confirmed stakeholders' concerns regarding access to the internet, noting that a barrier may be internet speed, in that a desired slot may be taken by another patient if the bandwidth is inadequate. In a systematic review, Zhao et al [21] concluded that patients' reluctance to adopt self-scheduling results from prior experience with the internet and computers, as well as preferences for communication methods. Addressing people's trust to enhance use is essential [29]. Researchers have identified gaps between people's interest in the technology and its use [29,44], and awareness of the technology and its use [71,75].

Cosmopolitanism—the extent to which an organization is networked with others external to itself—and peer pressure have not been discussed in the literature.

External Policy and Incentives

Research was influenced by government policies in several studies: a federally funded initiative was established to fast-track the advancement of health information technologies across Canada [36]. The British government recommended the *novel use of information technology* to meet government-mandated targets for appointment offerings [19]. The *Choose and Book System* studied by Parmar et al [34] was the national electronic referral and booking service introduced in England in 2004 which has since been replaced. Studies by researchers from China described the web-based appointment system, the use of which, as of 2009, has been supported by the Ministry of Health for deployment by all hospitals [24,48]. In Australia, the

National E-Health Strategy incorporated electronic communication between patients and providers [45]. Iran mandated that hospitals offer self-scheduling for outpatients, although compliance has been limited [47].

In their multinational research in Europe, Santana et al [44] acknowledged the importance of the prevailing legal and regulatory environment of each nation, as well as a country's health care policies and technological advances, in the adoption of self-scheduling. The influence of external policy and incentives at the national level on all aspects of eHealth have been scrutinized by researchers worldwide [77].

In addition to the impact of the government, other external factors may play a role in the uptake of self-scheduling including the COVID-19 pandemic [31].

Inner Setting

The key elements of the structural characteristics of the research settings are included in Multimedia Appendix 2. Of the 28 studies that defined the research setting, 14 (50%) were based in outpatient practices [5,18,19,32,34-37,43,45,70-72,74], 10 (36%) were based in medical centers [22, 24, 26, 28, 31, 42, 47, 48, 73, 75], and 4 (14%) surveyed community members [23,29,44,69]. Among the outpatient practice studies, 13% (4/30) featured settings of single specialties: 7% (2/30) dermatology [35,37], 3% (1/30) audiology [34], and 3% (1/30) genitourinary [19].

Data were not included in the studies for networks and communication or culture. Limited information was provided about the implementation climate. Friedman [18] conveyed that his physician colleagues "turned white as ghosts" at the suggestion of implementing self-scheduling, citing concerns about transparency; however, most adopted the platform. Acknowledging reluctance, Craig [33] advised, "like anything new, [self-scheduling] will take some getting used to."

Habibi et al [5] determined the importance of rendering favorable services owing to *increased competition*. This study was joined by 9 others that expressed the priority for change [18,22,24,26,28,31,42,45,47]. The sense of urgency increased over time. Zhang et al [24] reported lines forming late at night and "incidents of knife attacks at hospitals" resulting from patients' frustrations.

The importance of problem solving in the outpatient environment, which is the face of the hospital, was emphasized [26]. Lee et al [22] concluded that the impression of service quality put forth by the self-scheduling technology was a key success factor for a hospital to "gain an...advantage...in an increasingly competitive healthcare market." Volk et al [28] described the current environment that led to the introduction of the intervention as "threatening the organization's reputation and financial well-being."

Readiness for implementation was not addressed in detail: 3% (1/30) of studies [32] mentioned about providing the secretaries with a tablet and training; however, no other study described the engagement of leadership, available resources, or access to knowledge and information.

Characteristics of Individuals

Overview

Limited information in the body of literature included in this study was provided about individuals engaged in self-scheduling. In all, 3% (1/30) of studies described the hesitancy of physicians, although a revision to the intervention (pop-up menus) was developed during the project to address it [36]. Habibi et al [32] reflected on the "interest and eagerness of physicians," which contributed to the success of the self-scheduling intervention. The other articles in the scoping study offered little insight into the characteristics of the individuals participating in the intervention and whether individuals served as barriers to or facilitators of adoption.

Process

Limited information was provided about the process associated with the intervention: planning, engaging, executing, and reflecting and evaluating. Of the 30 studies, 1 (3%) study [36] elaborated on the importance of managing the physicians' expectations about slot availability, as patients may lose interest and discontinue the use of the system based on insufficient slots. In all, 7% (2/30) of studies postulated the importance of integrating the self-scheduling platform with the electronic medical record system [33,36]. Volk et al [28] documented a leadership task force. The literature offers limited insights into the implementation process.

Discussion

Existing Knowledge

This scoping review located 30 published articles that described synchronous, automated self-scheduling tools for patient appointments. The number of studies related to self-scheduling increased over time. The growing volume of research reflects the popularity of the technology, signaling its broadening appeal. Research performed in the same community-based clinic setting concluded a low intention to use [45,70,71]. However, low intention to use was not demonstrated in a study since 2015, perhaps reflecting the now pervasive use of computers. Patients' trust in the intervention has been studied as a possible barrier to the intervention [29]. Studies have continued to identify gaps between the interest and awareness of the technology and its use [29,75]. Researchers have concluded that concerns about access to the internet persist [26]. The introduction of self-scheduling in the context of a hospital as a business entity with financial interests commenced in 2020, perhaps reflecting the opportunity that a self-scheduling offering is no longer considered an initiative to appeal to innovators but rather a necessity of service delivery. Lee et al [22] determined that ease of use was no longer a factor of patients' continuous use, concluding that the system is now "stable, reliable, and well designed." This study reflected patients' increasing comfort with technology, which is supported by the literature about other consumer-oriented offerings such as telemedicine [78,79]. Articles aimed at optimization methods for scheduling, such as recommendations for demand matching [80], were formulated on a platform of automated scheduling, a reflection that the

literature has evolved from the foundational elements of implementation to a more sophisticated approach.

Efforts to determine the effect of self-scheduling may be hindered by the incorporation of the intervention as an element in a suite of technologies. Of the 30 studies, 11 (37%) studies in this scoping review [19,22,28,29,31,42,44,69,72,74,75] included self-scheduling as a component of a larger technology initiative, which may indicate that another intervention that was aligned with self-scheduling was the source of the organizational benefit.

The scoping study incorporated a systematic review that was conducted in 2017. The systematic review [21] reported the advantages of self-scheduling for organizations. In the literature before the systematic review, most gains were reported to have the potential to benefit the organization. Beginning in 2017, the advantages of self-scheduling have increasingly focused on the outer setting. Organizations react to consumers' access to technology and their competitive environment. Furthermore, the benefits of self-scheduling from the patients' perspective—satisfaction, time, convenience, and engagement—were increasingly referred to as potential rewards. Table 3 highlights the changes in the focus of the literature related to the identified need for the intervention. This may reflect an alteration in the determinants of adoption.

In a systematic review, Zhao et al [21] concluded that cost, flexibility, safety, and integrity were the barriers to adoption. Except for safety, these organizational barriers have not been replicated in the literature since 2017 [26,73]. However, the research upon which these conclusions were based drew upon the popular literature except for a 2004 case study [18] and a 2007 commentary [33], both of which noted providers' hesitancy. The lack of evidence-based organizational barriers over time may mean that the obstacles have historically been organizations' perceptions of patient behavior. The reluctance of patients to adopt based on their experience with computers reported in the systematic review [21] was not reproduced other than the potential impact of broadband speed noted in a focus group [26]. Despite the lack of evidence-based barriers, use of self-scheduling has continued to be reported at low rates during the period of 2017-2020 [47,72,75].

Table 3. Identified need for self-scheduled based on literature mentions.

Identified need	Mentions, n (%)	
	Before 2017 (n=23)	2017-2020 (n=25)
Inner setting	**14 (61)**	**9 (36)**
Organization's cost and labor	4 (17) [18,23,45,70]	5 (20) [28,29,37,73,75]
Organization's resource use (no-shows)	6 (26) [33-36,42,45]	3 (12) [29,37,73]
Organization's communication and information transparency	2 (9) [23,44]	1 (4) [22]
Alternative to organization's existing scheduling method	2 (9) [19,74]	0 (0)
Outer setting	**9 (39)**	**16 (64)**
Consumer access to technology	1 (4) [71]	1 (4) [32]
Organization's need to compete	1 (4) [42]	1 (4) [5]
Government policy	1 (4) [19]	1 (4) [47]
Patient satisfaction	1 (4) [42]	4 (16) [26,29,37,47]
Patient convenience	2 (9) [70,74]	5 (20) [31,43,69,72,73]
Patient wait time	3 (13) [23,24,48]	3 (12) [29,32,37]
Patient engagement	0 (0)	1 (4) [29]

Opportunities for Research

Self-scheduling may offer value to health care organizations. Additional research regarding the barriers to and facilitators of implementation is warranted.

Nomenclature

The terminology used to describe self-scheduling presented a challenge for the scoping study. The function—*scheduling*—was documented using a variety of labels, leading to a diversity of terms for the intervention under study. Standard terminology was not present in the research findings: the US-based research incorporated insurance coverage, lacking direct comparison with the non–US-based research that incorporated findings about *social grade* [71,72] and *socioeconomic status* [45]. Other characteristics, such as age range, varied in reporting. The lack of a standard vocabulary for the intervention and its users, uptake, evidence, and so forth has implications for research, as well as acceptance and adoption by health care organizations. This may present a barrier to organizations seeking knowledge about self-scheduling. Authors should incorporate keywords that reflect both breadth and depth to boost identification [81].

Implementation Framework

Within the CFIR, much of the research to date has focused on the intervention characteristics of self-scheduling, including the intervention source, relative advantage, adaptability, trialability, complexity, and design quality and packaging. The characteristics are largely presented as effects of the intervention, not the determinants of implementation. Evidence

strength and quality may be enhanced through improved research methods. The discussion of the cost of the intervention and its ongoing maintenance is limited. There is no consistent approach to the study of the intervention's characteristics to inform adoption. After presenting the results of a pilot study, researchers in 2020 [37] concluded the following:

> *We hope to encourage other colleagues to explore and share their experiences...and to stimulate conversation regarding implementation of technology to improve access to care.*

This request may signal a current gap in the literature regarding barriers to and facilitators of the implementation of self-scheduling.

Concepts warranting further research include the inner setting and individual characteristics contained in the CFIR. Qualitative research is needed to provide context and understanding of why health care organizations face barriers to successful outcomes identified by quantitative surveys. These may be present in the inner setting of organizations and individuals' characteristics, constructs that are largely unexplored by research on self-scheduling.

Although there is no consistent definition or inclusion of characteristics, within the outer setting, patient needs and resources in the form of gender, race, socioeconomic status, education level, employment, geography, computer access, experience, and literacy were explored by researchers. The nonstandard approach makes it difficult to determine the barriers to and facilitators of health care organizations to meet patients' needs. For example, rural populations face more problems in accessing care [82,83]. Consideration may be given to customized interventions for vulnerable patient populations, a topic unexplored in the literature. Otherwise, existing inequities related to the broadening gap of rural–urban disparities in life expectancy may be perpetuated [84].

External policy and incentives play a role in influencing self-scheduling, primarily at the country level. Although researchers mention the national initiatives, no details were provided about the initiative serving as a barrier or facilitator, or how that influence could be successful. Recognizing the importance of policies and regulations in health care technology [85], researchers may explore the characteristics and impact of external policies and incentives for nations that require self-scheduling to be offered by health care organizations.

Technology in Health Care

Researchers have explored the challenges of implementing other information and communication technologies that have exhibited evidence for improving systems, processes, and outcomes in health care. Documented inner setting obstacles to technology implementation include a culture that lacks receptivity [86], an absence of trust [87], a resistance to change [88], workflow changes that were required for uptake [89,90], and upfront and ongoing costs of the solution [91]. The Systems Engineering Initiative for Patient Safety 2.0 model was introduced to account for human factors systems, extending into the concepts of adaptation, engagement, and configuration [92]. The determinants identified by researchers evaluating the implementation of other technologies by health care organizations may offer insight into a framework to explore the limited uptake of self-scheduling.

Health Care Providers

Although there are references to the providers' perspective in the academic literature incorporated in this scoping study [18,32,33,36], these have not been examined in detail. For the only study that reported measuring it, physician punctuality improved after the intervention was introduced, and the researchers surmised that the enhancement resulted from the physicians' enthusiasm about the solution, as well as the reminder of the first appointment of the day transmitted via text from the self-scheduling tool [32]. Although 3% (1/30) of studies [26] concluded that they were able to eliminate some elements of patient dissatisfaction, the researchers determined that 40% of the dissatisfaction was a function of the physicians being late and canceling clinics, albeit the intervention they launched enabled the staff to inform patients of the delays. The connectivity of the intervention to its offering—the provider's time—is largely unexplored.

To date, the literature on the uptake of self-scheduling has focused on the end user: patients' awareness, characteristics, use, and intention to use. As self-scheduling platforms aim to provide a limited inventory of providers' time, the provider is an equally important stakeholder. Further research may reveal ideas, variables, and determinants that are not yet recognized by health care organizations. The literature needs to focus more on the integration of technology into work systems. Research on providers as resisters of other automated health care administrative tools, such as telemedicine, has proliferated [93]. Similar research techniques may be applied to garner a better understanding of self-scheduling.

Relationships

The existing literature does not elucidate the factors that promote or impede the uptake of self-scheduling by health care organizations. The absence of aggregation and examination of barriers and facilitators may reflect the complexity of self-scheduling as an intervention. As demonstrated in the literature, the solution is influenced by the intervention's characteristics, the outer and inner settings of the health care organization, individual stakeholders, and the process related to the intervention. Self-scheduling cannot be implemented and scaled without a comprehensive understanding of these factors. In contrast to the focus on dissecting individual components defined by the CFIR, the success of an implementation by a complex, adaptive health care organization is informed by the interdependence of the determinants [94]. The exploration of enablers and obstacles by examining the contingent and reciprocal relationships within health care organizations may better illuminate the implementation determinants for self-scheduling.

Limitations

The author (EW) conducted the screening process, which may have introduced selection bias. The lack of a standard naming convention may have resulted in missing relevant articles for the scoping review. Given the large number of findings from

countries with a primary language other than English, the inclusion of English-only articles may have missed publications that were not accessible from the databases deployed in the search strategy.

In contrast to systemic reviews, scoping studies, by definition, do not incorporate a quality assessment of individual studies; therefore, it is challenging to assess whether studies produce robust findings [64]. As such, data synthesis and interpretation are limited [63].

An agreement on common measures to identify and monitor the impact of self-scheduling is required. Research that tracked the most cited advantage of reducing the no-show rate failed to accompany the discourse with a definition of said rate.

Conclusions

This scoping review cataloged existing knowledge and identified gaps in knowledge regarding the uptake of automated self-scheduling by health care organizations. The intervention was defined. There was evidence of the broadening appeal and demonstrable benefits of automated self-scheduling; however, the uptake remained low.

Prior research examined implementation effectiveness; this review focused on barriers to and facilitators of self-scheduling by health care organizations. Outer setting determinants to include national policy, competition, the response to patients' needs, and technology access played an increasing role in influencing implementation over time. Automated self-scheduling may offer a solution to health care organizations striving to positively affect access.

Acknowledgments

The author acknowledges the contributions of Dr Doug Hough, Dr Kathy McDonald, Dr Michael Rosen, Dr Aditi Sen, Dr Jonathan Weiner, Dr Christina Yuan, and Ms Claire Twose of Johns Hopkins University.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[DOCX File , 16 KB - jmir_v24i1e28323_app1.docx]

Multimedia Appendix 2
Charted data.
[XLSX File (Microsoft Excel File), 19 KB - jmir_v24i1e28323_app2.xlsx]

References

1. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington, DC: National Academies Press; 2001.
2. Institute of Medicine. Transforming Health Care Scheduling and Access: Getting to Now. Washington, DC: The National Academies Press; 2015.
3. Leddy KM, Kaldenberg DO, Becker BW. Timeliness in ambulatory care treatment. An examination of patient satisfaction and wait times in medical practices and outpatient test and treatment facilities. J Ambul Care Manage 2003;26(2):138-149. [doi: 10.1097/00004479-200304000-00006] [Medline: 12698928]
4. Kerwin KE. The role of the internet in improving healthcare quality. J Healthc Manag 2002;47(4):225-236. [doi: 10.1097/00115514-200207000-00005]
5. Habibi M, Abadi F, Tabesh H, Vakili-Arki H, Abu-Hanna A, Eslami S. Evaluation of patient satisfaction of the status of appointment scheduling systems in outpatient clinics: identifying patients' needs. J Adv Pharm Technol Res 2018;9(2):51-55. [doi: 10.4103/japtr.japtr_134_18]
6. Waller J, Jackowska M, Marlow L, Wardle J. Exploring age differences in reasons for nonattendance for cervical screening: a qualitative study. BJOG 2012 Jan;119(1):26-32. [doi: 10.1111/j.1471-0528.2011.03030.x] [Medline: 21668764]
7. Murray M, Berwick DM. Advanced access: reducing waiting and delays in primary care. J Am Med Assoc 2003 Feb 26;289(8):1035-1040. [doi: 10.1001/jama.289.8.1035] [Medline: 12597760]
8. Rust G, Ye J, Baltrus P, Daniels E, Adesunloye B, Fryer G. Practical barriers to timely primary care access: impact on adult use of emergency department services. Arch Intern Med 2008 Aug 11;168(15):1705-1710. [doi: 10.1001/archinte.168.15.1705] [Medline: 18695087]
9. Bhandari G, Snowdon A. Design of a patient-centric, service-oriented health care navigation system for a local health integration network. Behav Inform Technol 2012 Mar;31(3):275-285. [doi: 10.1080/0144929x.2011.563798]

10. Schneider EC, Zaslavsky AM, Landon BE, Lied TR, Sheingold S, Cleary PD. National quality monitoring of Medicare health plans: the relationship between enrollees' reports and the quality of clinical care. Med Care 2001 Dec;39(12):1313-1325. [doi: 10.1097/00005650-200112000-00007] [Medline: 11717573]
11. Glogovac G, Kennedy M, Weisgerber M, Kakazu R, Grawe B. Wait times in musculoskeletal patients: what contributes to patient satisfaction. J Patient Exp 2020 Aug;7(4):549-553 [FREE Full text] [doi: 10.1177/2374373519864828] [Medline: 33062877]
12. Marco CA, Bryant M, Landrum B, Drerup B, Weeman M. Refusal of emergency medical care: an analysis of patients who left without being seen, eloped, and left against medical advice. Am J Emerg Med 2021 Feb;40:115-119. [doi: 10.1016/j.ajem.2019.158490] [Medline: 31704062]
13. Jansson B. Choosing a good appointment system—A study of queues of the type (D, M, 1). Oper Res 1966 Apr;14(2):292-312. [doi: 10.1287/opre.14.2.292]
14. Mak H, Rong Y, Zhang J. Sequencing appointments for service systems using inventory approximations. Manuf Serv Oper Manag 2014 May;16(2):251-262. [doi: 10.1287/msom.2013.0470]
15. Shugan SM, Xie J. Advance pricing of services and other implications of separating purchase and consumption. J Serv Res 2016 Jun 29;2(3):227-239. [doi: 10.1177/109467050023001]
16. Chiu C. Understanding relationship quality and online purchase intention in e-tourism: a qualitative application. Qual Quant 2007 Nov 23;43(4):669-675. [doi: 10.1007/s11135-007-9147-6]
17. Robinson LW, Chen RR. Estimating the implied value of the customer's waiting time. Manuf Serv Oper Manag 2011 Jan;13(1):53-57. [doi: 10.1287/msom.1100.0304]
18. Friedman JP. Internet patient scheduling in real-life practice. J Med Pract Manage 2004;20(1):13-15. [Medline: 15500015]
19. Jones R, Menon-Johansson A, Waters AM, Sullivan AK. eTriage--a novel, web-based triage and booking service: enabling timely access to sexual health clinics. Int J STD AIDS 2010 Jan 01;21(1):30-33. [doi: 10.1258/ijsa.2008.008466] [Medline: 19884355]
20. Idowu P, Adeosun O, Williams K. Dependable online appointment booking system for NHIS outpatient in Nigerian teaching hospitals. Int J Comput Sci Inform Technol 2014 Aug 31;6(4):59-73. [doi: 10.5121/ijcsit.2014.6405]
21. Zhao P, Yoo I, Lavoie J, Lavoie BJ, Simoes E. Web-based medical appointment systems: a systematic review. J Med Internet Res 2017 Apr 26;19(4):e134 [FREE Full text] [doi: 10.2196/jmir.6747] [Medline: 28446422]
22. Lee Y, Tsai H, Ruangkanjanases A. The determinants for food safety push notifications on continuance intention in an e-appointment system for public health medical services: the perspectives of UTAUT and information system quality. Int J Environ Res Public Health 2020 Nov 09;17(21):8287 [FREE Full text] [doi: 10.3390/ijerph17218287] [Medline: 33182479]
23. Chen S, Liu S, Li S, Yen DC. Understanding the mediating effects of relationship quality on technology acceptance: an empirical study of e-appointment system. J Med Syst 2013 Dec 19;37(6):9981. [doi: 10.1007/s10916-013-9981-0] [Medline: 24141491]
24. Zhang M, Zhang C, Sun Q, Cai Q, Yang H, Zhang Y. Questionnaire survey about use of an online appointment booking system in one large tertiary public hospital outpatient service center in China. BMC Med Inform Decis Mak 2014 Jun 09;14:49 [FREE Full text] [doi: 10.1186/1472-6947-14-49] [Medline: 24912568]
25. Kamo N, Bender AJ, Kalmady K, Blackmore CC. Meaningful use of the electronic patient portal - Virginia Mason's journey to create the perfect online patient experience. Healthc (Amst) 2017 Dec;5(4):221-226. [doi: 10.1016/j.hjdsi.2016.09.003] [Medline: 27727028]
26. Mendoza S, Padpad R, Vael A, Alcazar C, Pula R. A web-based “InstaSked” appointment scheduling system at perpetual help medical center outpatient department. In: EAI/Springer Innovations in Communication and Computing. Cham: Springer; 2020:3-14.
27. Gupta D, Denton B. Appointment scheduling in health care: challenges and opportunities. IIE Transactions 2008 Jul 21;40(9):800-819. [doi: 10.1080/07408170802165880]
28. Volk A, Davis M, Abu-Ghname A, Warfield R, Ibrahim R, Karon G, et al. Ambulatory access: improving scheduling increases patient satisfaction and revenue. Plast Reconstr Surg 2020 Oct;146(4):913-919. [doi: 10.1097/PRS.0000000000007195] [Medline: 32970013]
29. Xie H, Prybutok G, Peng X, Prybutok V. Determinants of trust in health information technology: an empirical investigation in the context of an online clinic appointment system. Int J Hum–Comput Interact 2020 Jan 14;36(12):1095-1109. [doi: 10.1080/10447318.2020.1712061]
30. Sherly I, Mahalakshmi A, Menaka D, Sujatha R. Online appointment reservation and scheduling for healthcare-a detailed study. Int J Innov Res Comput Commun Eng 2016;4(2):2053-2059. [doi: 10.15680/IJIRCCE.2016. 0402056]
31. Judson T, Odisho A, Neinstein A, Chao J, Williams A, Miller C, et al. Rapid design and implementation of an integrated patient self-triage and self-scheduling tool for COVID-19. J Am Med Inform Assoc 2020 Jun 01;27(6):860-866 [FREE Full text] [doi: 10.1093/jamia/ocaa051] [Medline: 32267928]
32. Habibi MR, Mohammadabadi F, Tabesh H, Vakili-Arki H, Abu-Hanna A, Eslami S. Effect of an online appointment scheduling system on evaluation metrics of outpatient scheduling system: a before-after MulticenterStudy. J Med Syst 2019 Jul 12;43(8):281. [doi: 10.1007/s10916-019-1383-5] [Medline: 31300894]
33. Craig A. Self-service appointments. Adv Nurse Pract 2007 Oct;15(10):24-25. [Medline: 19998876]

34. Parmar V, Large A, Madden C, Das V. The online outpatient booking system 'Choose and Book' improves attendance rates at an audiology clinic: a comparative audit. Inform Prim Care 2009 Sep 01;17(3):183-186 [FREE Full text] [doi: 10.14236/jhi.v17i3.733] [Medline: 20074431]
35. Siddiqui Z, Rashid R. Cancellations and patient access to physicians: ZocDoc and the evolution of e-medicine. Dermatol Online J 2013;19(4):14. [doi: 10.5070/D358k4d6wp]
36. Paré G, Trudel M, Forget P. Adoption, use, and impact of e-booking in private medical practices: mixed-methods evaluation of a two-year showcase project in Canada. JMIR Med Inform 2014 Sep 24;2(2):e24 [FREE Full text] [doi: 10.2196/medinform.3669] [Medline: 25600414]
37. Yanovsky R, Das S. Patient-initiated online appointment scheduling: pilot program at an urban academic dermatology practice. J Am Acad Dermatol 2020 Nov;83(5):1479-1481. [doi: 10.1016/j.jaad.2020.03.035] [Medline: 32213308]
38. Gao J, Wong T, Wang C. Coordinating patient preferences through automated negotiation: a multiagent systems model for diagnostic services scheduling. Adv Eng Inform 2019 Oct;42:100934. [doi: 10.1016/j.aei.2019.100934]
39. Tang J, Yan C, Cao P. Appointment scheduling algorithm considering routine and urgent patients. Expert Syst Appl 2014 Aug;41(10):4529-4541. [doi: 10.1016/j.eswa.2014.01.014]
40. Tuzovic S, Kuppelwieser V. Developing a framework of service convenience in health care: an exploratory study for a primary care provider. Health Mark Q 2016;33(2):127-148. [doi: 10.1080/07359683.2016.1166840] [Medline: 27215644]
41. Berry LL, Beckham D, Dettman A, Mead R. Toward a strategy of patient-centered access to primary care. Mayo Clin Proc 2014 Oct;89(10):1406-1415. [doi: 10.1016/j.mayocp.2014.06.011] [Medline: 25199953]
42. Chang HH, Chang CS. An assessment of technology-based service encounters and network security on the e-health care systems of medical centers in Taiwan. BMC Health Serv Res 2008 Apr 17;8(1):87 [FREE Full text] [doi: 10.1186/1472-6963-8-87] [Medline: 18419820]
43. Kurtzman GW, Keshav MA, Satish NP, Patel MS. Scheduling primary care appointments online: differences in availability based on health insurance. Healthc (Amst) 2018 Sep;6(3):186-190. [doi: 10.1016/j.hjdsi.2017.07.002] [Medline: 28757308]
44. Santana S, Lausen B, Bujnowska-Fedak M, Chronaki C, Kummervold PE, Rasmussen J, et al. Online communication between doctors and patients in Europe: status and perspectives. J Med Internet Res 2010;12(2):e20 [FREE Full text] [doi: 10.2196/jmir.1281] [Medline: 20551011]
45. Zhang X, Yu P, Yan J, Ton AM. Using diffusion of innovation theory to understand the factors impacting patient acceptance and use of consumer e-health innovations: a case study in a primary care clinic. BMC Health Serv Res 2015 Feb 21;15:71 [FREE Full text] [doi: 10.1186/s12913-015-0726-2] [Medline: 25885110]
46. Wali P, Salim M, Ahmed BI. A prototype mobile application for clinic appointment reminder and scheduling system in Erbil city. Int J Adv Sci Technol 2019;28(1):17-24 [FREE Full text]
47. Samadbeik M, Saremian M, Garavand A, Hasanvandi N, Sanaeinasab S, Tahmasebi H. Assessing the online outpatient booking system. Shiraz E-Med J 2018 Mar 11:e60249. [doi: 10.5812/semj.60249]
48. Cao W, Wan Y, Tu H, Shang F, Liu D, Tan Z, et al. A web-based appointment system to reduce waiting for outpatients: a retrospective study. BMC Health Serv Res 2011 Nov 22;11:318 [FREE Full text] [doi: 10.1186/1472-6963-11-318] [Medline: 22108389]
49. Yu W, Yu X, Hu H, Duan G, Liu Z, Wang Y. Use of hospital appointment registration systems in China: a survey study. Glob J Health Sci 2013 Jul 22;5(5):193-201 [FREE Full text] [doi: 10.5539/gjhs.v5n5p193] [Medline: 23985121]
50. Yeh M, Saltman RB. Creating online personal medical accounts: recent experience in two developed countries. Health Policy Technol 2019 Jun;8(2):171-178. [doi: 10.1016/j.hlpt.2019.05.004]
51. Kruk ME, Gage AD, Arsenault C, Jordan K, Leslie HH, Roder-DeWan S, et al. High-quality health systems in the Sustainable Development Goals era: time for a revolution. Lancet Glob Health 2018 Nov;6(11):1196-1252. [doi: 10.1016/S2214-109X(18)30386-3]
52. Malhotra C, Do YK. Socio-economic disparities in health system responsiveness in India. Health Policy Plan 2013 Mar;28(2):197-205 [FREE Full text] [doi: 10.1093/heapol/czs051] [Medline: 22709921]
53. Ghafur S, Schneider E. Why are health care organizations slow to adopt patient-facing digital technologies? Health Affairs. 2019. URL: https://www.healthaffairs.org/do/10.1377/hblog20190301.476734/full/ [accessed 2021-02-14]
54. Bernstein ML, McCreless T, Côté MJ. Five constants of information technology adoption in healthcare. Hospital Topics 2007 Jan;85(1):17-25. [doi: 10.3200/htps.85.1.17-26]
55. Ingebrigtsen T, Georgiou A, Clay-Williams R, Magrabi F, Hordern A, Prgomet M, et al. The impact of clinical leadership on health information technology adoption: systematic review. Int J Med Inform 2014 Jun;83(6):393-405. [doi: 10.1016/j.ijmedinf.2014.02.005] [Medline: 24656180]
56. Jacob C, Sanchez-Vazquez A, Ivory C. Social, organizational, and technological factors impacting clinicians' adoption of mobile health tools: systematic literature review. JMIR Mhealth Uhealth 2020 Feb 20;8(2):e15935 [FREE Full text] [doi: 10.2196/15935] [Medline: 32130167]
57. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]

58. Versel N. Online reservations: letting patients make their own appointments. AMEDNEWS. 2004. URL: http://www.providersedge.com/ehdocs/ehr_articles/online_reservations-letting_patients_make_their_own_appointments.pdf [accessed 2020-12-31]
59. Wang W, Gupta D. Adaptive appointment systems with patient preferences. Manuf Serv Oper Manag 2011 Jul;13(3):373-389. [doi: 10.1287/msom.1110.0332]
60. Lowes R. Let patients book their own appointments? Med Econ 2006 Jun 02;83(11):27-28. [Medline: 16827528]
61. Riddell S. Technology: the race to get the NHS online. Health Serv J 2012 Sep 06;122(6319):32. [Medline: 23074883]
62. Farr C. Dentistry takes the cybercure: scheduling, consultations, records move to the net. Dent Today 2000 May;19(5):106-113. [Medline: 12524769]
63. Armstrong R, Hall B, Doyle J, Waters E. Cochrane update. 'Scoping the scope' of a cochrane review. J Public Health (Oxf) 2011 Mar;33(1):147-150 [FREE Full text] [doi: 10.1093/pubmed/fdr015] [Medline: 21345890]
64. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
65. Rethlefsen ML, Kirtley S, Waffenschmidt S, Ayala AP, Moher D, Page MJ, PRISMA-S Group. PRISMA-S: an extension to the PRISMA statement for reporting literature searches in systematic reviews. Syst Rev 2021 Jan 26;10(1):39 [FREE Full text] [doi: 10.1186/s13643-020-01542-z] [Medline: 33499930]
66. Stone AB, Yuan CT, Rosen MA, Grant MC, Benishek LE, Hanahan E, et al. Barriers to and facilitators of implementing enhanced recovery pathways using an implementation framework: a systematic review. JAMA Surg 2018 Mar 01;153(3):270-279. [doi: 10.1001/jamasurg.2017.5565] [Medline: 29344622]
67. Chung S, Martinez M, Frosch D, Jones V, Chan A. Patient-centric scheduling with the implementation of health information technology to improve the patient experience and access to care: retrospective case-control analysis. J Med Internet Res 2020 Jun 10;22(6):e16451 [FREE Full text] [doi: 10.2196/16451] [Medline: 32519970]
68. Zhang MW, Ho RC, Hawa R, Sockalingam S. Pilot implementation and user preferences of a Bariatric After-care application. Technol Health Care 2015 Oct 27;23(6):729-736. [doi: 10.3233/thc-151025]
69. Ryan M, Waller J, Marlow LA. Could changing invitation and booking processes help women translate their cervical screening intentions into action? A population-based survey of women's preferences in Great Britain. BMJ Open 2019 Jul 11;9(7):e028134 [FREE Full text] [doi: 10.1136/bmjopen-2018-028134] [Medline: 31300499]
70. Zhang X, Yu P, Yan J. Patients' adoption of the e-appointment scheduling service: a case study in primary healthcare. Stud Health Technol Inform 2014;204:176-181. [Medline: 25087546]
71. Zhang X, Yu P, Yan J, Hu H, Goureia N. Patients' perceptions of web self-service applications in primary healthcare. Stud Health Technol Inform 2012;178:242-249. [Medline: 22797048]
72. Ryan M, Marlow L, Forster A, Ruwende J, Waller J. Offering an app to book cervical screening appointments: a service evaluation. J Med Screen 2020 Jun 09;27(2):85-89 [FREE Full text] [doi: 10.1177/0969141319871312] [Medline: 31500520]
73. Ganguli I, Orav EJ, Lupo C, Metlay JP, Sequist TD. Patient and visit characteristics associated with use of direct scheduling in primary care practices. JAMA Netw Open 2020 Aug 03;3(8):e209637 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.9637] [Medline: 32852551]
74. Denizard-Thompson NM, Feiereisel KB, Stevens SF, Miller DP, Wofford JL. The digital divide at an urban community health center: implications for quality improvement and health care access. J Community Health 2011 Jun 11;36(3):456-460. [doi: 10.1007/s10900-010-9327-5] [Medline: 21086028]
75. Chen P, Xiao L, Gou Z, Xiang L, Zhang X, Feng P. Telehealth attitudes and use among medical professionals, medical students and patients in China: a cross-sectional survey. Int J Med Inform 2017 Dec;108:13-21. [doi: 10.1016/j.ijmedinf.2017.09.009] [Medline: 29132618]
76. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009 Aug 07;4(1):50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
77. Friedman D. Assessing the potential of national strategies for electronic health records for population health monitoring and research. Vital Health Stat 2 2006 Jan(143):1-83 [FREE Full text] [Medline: 17552126]
78. Rimmer RA, Christopher V, Falck A, de Azevedo Pribitkin E, Curry JM, Luginbuhl AJ, et al. Telemedicine in otolaryngology outpatient setting-single center head and neck surgery experience. Laryngoscope 2018 Sep 15;128(9):2072-2075. [doi: 10.1002/lary.27123] [Medline: 29446446]
79. Andino JJ, Guduguntla V, Weizer A, Roberts WW, Wittmann D, Miller D, et al. Examining the value of video visits to patients in an outpatient urology clinic. Urology 2017 Dec;110:31-35. [doi: 10.1016/j.urology.2017.07.050] [Medline: 28843776]
80. Wen H, Song J, Pan X. Physician recommendation on healthcare appointment platforms considering patient choice. IEEE Trans Automat Sci Eng 2020 Apr;17(2):886-899. [doi: 10.1109/tase.2019.2950724]
81. Guise J, Chang C, Viswanathan M, Glick S, Treadwell J, Umscheid CA, et al. Agency for healthcare research and quality evidence-based practice center methods for systematically reviewing complex multicomponent health care interventions. J Clin Epidemiol 2014 Nov;67(11):1181-1191. [doi: 10.1016/j.jclinepi.2014.06.010] [Medline: 25438663]

82. Laditka JN, Laditka SB, Probst JC. Health care access in rural areas: evidence that hospitalization for ambulatory care-sensitive conditions in the United States may increase with the level of rurality. Health Place 2009 Sep;15(3):731-740. [doi: 10.1016/j.healthplace.2008.12.007] [Medline: 19211295]
83. Probst JC, Moore CG, Glover SH, Samuels ME. Person and place: the compounding effects of race/ethnicity and rurality on health. Am J Public Health 2004 Oct;94(10):1695-1703. [doi: 10.2105/ajph.94.10.1695] [Medline: 15451735]
84. Singh GK, Siahpush M. Widening rural-urban disparities in life expectancy, U.S., 1969-2009. Am J Prev Med 2014 Feb;46(2):19-29. [doi: 10.1016/j.amepre.2013.10.017] [Medline: 24439358]
85. Lang A. Government capacities and stakeholders: what facilitates ehealth legislation? Global Health 2014 Jan 13;10:4 [FREE Full text] [doi: 10.1186/1744-8603-10-4] [Medline: 24410989]
86. Police R, Foster T, Wong K. Adoption and use of health information technology in physician practice organisations: systematic review. Inform Prim Care 2010 Dec 01;18(4):245-258 [FREE Full text] [doi: 10.14236/jhi.v18i4.780] [Medline: 22040851]
87. Schreiweis B, Pobiruchin M, Strotbaum V, Suleder J, Wiesner M, Bergh B. Barriers and facilitators to the implementation of eHealth services: systematic literature analysis. J Med Internet Res 2019 Nov 22;21(11):e14197 [FREE Full text] [doi: 10.2196/14197] [Medline: 31755869]
88. Kruse CS, Kristof C, Jones B, Mitchell E, Martinez A. Barriers to electronic health record adoption: a systematic literature review. J Med Syst 2016 Dec;40(12):252 [FREE Full text] [doi: 10.1007/s10916-016-0628-9] [Medline: 27714560]
89. Dutta B, Hwang H. The adoption of electronic medical record by physicians: a PRISMA-compliant systematic review. Medicine (Baltimore) 2020 Feb;99(8):e19290 [FREE Full text] [doi: 10.1097/MD.0000000000019290] [Medline: 32080145]
90. Castillo VH, Martínez-García AI, Pulido JR. A knowledge-based taxonomy of critical factors for adopting electronic health record systems by physicians: a systematic literature review. BMC Med Inform Decis Mak 2010 Oct;10:60 [FREE Full text] [doi: 10.1186/1472-6947-10-60] [Medline: 20950458]
91. O'Donnell A, Kaner E, Shaw C, Haighton C. Primary care physicians' attitudes to the adoption of electronic medical records: a systematic review and evidence synthesis using the clinical adoption framework. BMC Med Inform Decis Mak 2018 Nov 13;18(1):101 [FREE Full text] [doi: 10.1186/s12911-018-0703-x] [Medline: 30424758]
92. Holden RJ, Carayon P, Gurses AP, Hoonakker P, Hundt AS, Ozok AA, et al. SEIPS 2.0: a human factors framework for studying and improving the work of healthcare professionals and patients. Ergonomics 2013 Nov;56(11):1669-1686 [FREE Full text] [doi: 10.1080/00140139.2013.838643] [Medline: 24088063]
93. Choi WS, Park J, Choi JY, Yang J. Stakeholders' resistance to telemedicine with focus on physicians: utilizing the Delphi technique. J Telemed Telecare 2018 Jan 01:378-385. [doi: 10.1177/1357633X18775853] [Medline: 29792080]
94. Sarkies M, Long JC, Pomare C, Wu W, Clay-Williams R, Nguyen HM, et al. Avoiding unnecessary hospitalisation for patients with chronic conditions: a systematic review of implementation determinants for hospital avoidance programmes. Implement Sci 2020 Oct 21;15(1):91 [FREE Full text] [doi: 10.1186/s13012-020-01049-0] [Medline: 33087147]

Abbreviations

CFIR: Consolidated Framework for Implementation Research
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-S: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Search

Edited by A Mavragani; submitted 01.03.21; peer-reviewed by P Zhao, A O'Donnell; comments to author 31.03.21; revised version received 25.04.21; accepted 26.11.21; published 11.01.22.

Please cite as:
Woodcock EW
Barriers to and Facilitators of Automated Patient Self-scheduling for Health Care Organizations: Scoping Review
J Med Internet Res 2022;24(1):e28323
URL: https://www.jmir.org/2022/1/e28323
doi:10.2196/28323
PMID:35014968

Review

Digital Storytelling for Health-Related Outcomes in Older Adults: Systematic Review

Jennifer Stargatt[1], BA(Hons); Sunil Bhar[1], PhD; Jahar Bhowmik[2], PhD; Abdullah Al Mahmud[3], PhD

[1]Department of Psychological Sciences, School of Health Sciences, Swinburne University of Technology, Hawthorn, Australia

[2]Department of Health Sciences and Biostatistics, School of Health Sciences, Swinburne University of Technology, Hawthorn, Australia

[3]Centre for Design Innovation, School of Design, Swinburne University of Technology, Hawthorn, Australia

Corresponding Author:
Jennifer Stargatt, BA(Hons)
Department of Psychological Sciences
School of Health Sciences
Swinburne University of Technology
John St
Hawthorn, 3122
Australia
Phone: 61 432489829
Email: jstargatt@swin.edu.au

Abstract

Background: Older adults face a unique set of challenges and may experience a range of psychological comorbidities. Digital storytelling is an emerging tool for sharing and recording lived experiences and may have the potential to support well-being but is yet to be systematically reviewed for use among older adults.

Objective: The aim of this review is to examine the methods for creating digital stories, the health-related outcomes associated with creating digital stories, and the potential for implementing digital storytelling with older adults.

Methods: We systematically searched electronic databases to identify articles published in English that reported on at least one health-related outcome of digital storytelling for participants aged ≥60 years. Data were extracted and synthesized using qualitative content analysis and summarized in tables. The methodological quality of the studies was assessed using the Mixed Methods Appraisal Tool.

Results: A total of 8 studies were included in the review. Participants were primarily community-dwelling older adults living with dementia, involving family caregivers and professional care staff. Studies have taken various approaches to digital storytelling and reported diverse benefits associated with digital storytelling, including improvements in mood, memory, social engagement, and quality of relationships. Although the potential for implementation was not widely examined, some studies have presented evidence for acceptability and feasibility. Generally, studies were of high quality, despite the absence of comparator groups and confounder analyses.

Conclusions: The evidence reviewed suggests that despite the various approaches taken, digital storytelling shows promise as an effective approach for supporting well-being in older adults.

Trial Registration: PROSPERO International Prospective Register of Systematic Reviews CRD42019145922; https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42019145922

International Registered Report Identifier (IRRID): RR2-10.2196/15512

(J Med Internet Res 2022;24(1):e28113) doi:10.2196/28113

KEYWORDS

digital storytelling; mental health; aging; dementia; reminiscence; memory; systematic review; older adults

Introduction

Background

The number of people aged ≥60 years is growing rapidly, worldwide [1]. Although most older adults experience positive mental health, a significant number of older adults experience psychological comorbidities, including depression, anxiety, and loneliness [2-4]. Furthermore, for those who require care, moving into a long-term care setting can elicit feelings of reduced personal autonomy, purpose, and sense of self, as familiar possessions and activities that support the person's identity are often lost [5,6].

Autobiographical storytelling may improve psychological well-being in older adults living in community or long-term care settings. Recalling personal stories may encourage beliefs of self-mastery and problem solving, improve mood by eliciting pleasant memories, and support ego integrity—accepting and integrating one's highs and lows and finding a meaning or greater purpose in life events [7,8]. Storytelling may enable older adults to feel recognized, affirmed, empowered, and accomplished and may assist in building resilience [8-10]. Reviews of studies have suggested that activities that include reminiscence about one's life story may improve subjective psychological well-being, quality of life, mood, and cognition [11-14]. In addition, reviews have found that activities that involve reminiscence have the potential to improve quality of life, cognition, communication, and activities of daily living in older adults living with Alzheimer disease and dementia [15,16].

Tangible artifacts, such as books, collages, and memory boxes, are often created as a product of such life story work with older adults to record, retain, and share stories with others [17]. An integrative review found that life story work, in which an end product was created, assisted in maintaining a sense of identity and enhancing relationships for older adults in long-term care settings, most of whom were living with dementia [18]. For older people living with dementia, life story work can stimulate memories, enhance person-centered care, and promote conversations with family members and carers [17,19,20].

Owing to the advances in the capability and accessibility of multimedia technologies, it is now possible to produce digital story artifacts with relative ease. Digital storytelling is a process that involves using multimedia technology to combine images, sounds, and narration to create a film that documents one's lived experiences [21]. It is an interdisciplinary approach used in educational settings [22,23], participatory research [24,25], and community engagement [26]. It can be facilitated in groups or one-on-one with individuals, with common aims to engage participants to record and share their lived experiences to educate others [27], enhance community engagement [28], and deepen their understanding of their personal stories [29]. For example, Lambert [21,30] engages people to create films about their own lived experiences, in which each story lasted for 3 to 5 minutes.

The use of digital storytelling, broadly defined, to improve the health of older adults is an emerging area of research. Similar to traditional life story artifacts, using digital technology to create and share autobiographical stories may enable older adults to benefit from the experience of being listened to and the opportunity to express their emotions and their identity [31]. To date, studies suggest that such digital storytelling is used with older adults, including those living with dementia, in a variety of ways—as a tool to improve mood [32,33], enhance memory [33], increase social connectedness [32,34,35], enhance the quality of care [34], and promote intergenerational relationships and learning [36-38].

With the increased accessibility of digital technologies, stories about past experiences can be easily documented in the form of narratives. However, the methods for creating such stories, the outcomes of such stories for personal well-being, and the potential to implement digital story activities within community or long-term care settings remain to be systematically reviewed. Given that digital storytelling could be beneficial for older adults, a systematic exploration of this growing body of literature is warranted.

Objectives

This review aims to answer the following questions:

1. What health-related outcomes have been reported in relation to digital storytelling activities in older adults?
2. What methods for conducting digital storytelling activities for older adults have been reported?
3. What is the potential for implementation (eg, acceptability, appropriateness, and feasibility) of digital storytelling activities for older adults?

Methods

Registration

The peer-reviewed systematic review protocol [39] was developed following the PRISMA-P (Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols) guidelines [40] and registered with PROSPERO (CRD42019145922). This systematic review adhered to the recommendations of the PRISMA (Preferred Reporting Items for Systematic Review and Meta-Analyses) guidelines [41].

Eligibility Criteria

Study Designs

In this review, we included all possible study designs, including quantitative (eg, randomized, nonrandomized, quasi-randomized, and cluster-randomized controlled trials; pilot trials; open trials; case studies; cross-section studies; cohort studies; and case-control studies), qualitative, and mixed methods studies, provided that at least one health-related outcome was reported concerning digital storytelling. No study designs were excluded provided all other inclusion criteria were met. This decision to include a range of study designs was pragmatic, given that digital storytelling remains a relatively new area of research across various health care disciplines. Hence, the breadth of studies would provide an overview of the methods, outcomes, and implementation characteristics of digital storytelling with older adults.

Participants

We included studies in which all participants were older adults, defined for this review by the United Nations classification of those ≥60 years [1]. No exclusions were made based on participant health—studies were included regardless of dementia, mild cognitive impairment (MCI), and other illnesses. Studies were included regardless of participant setting (eg, community, long-term care, and hospitals).

Interventions

Digital storytelling was defined as creating a short (usually 3-5 minutes) multimedia clip (eg, images, videos, narration, and music) focused on the lived experience of older adults. We did not exclude studies based on story length or the level of participant involvement in production; digital stories may have been produced entirely by participants, produced on their behalf, or cocreated by the participants and others such as researchers, carers, or volunteers.

Comparator Groups

Studies were included regardless of whether they had a comparator group and irrespective of the type of comparator group included.

Outcomes

Health-Related Outcomes

Studies examining any outcome related to physical, psychological, or social health were included in the review. Examples of such outcomes include mood, memory, quality of life, and social engagement. Studies were included irrespective of whether these outcomes were measured quantitatively (eg, using validated psychometric assessment tools) or assessed qualitatively (eg, as a result of participant interviews, which were transcribed and analyzed thematically). Studies were excluded if digital storytelling was used in conjunction with another activity where the effects of digital storytelling alone were not reported or could not be ascertained.

Methods of Storytelling

The outcomes related to methods used in digital storytelling that were reviewed were as follows: (1) process characteristics (duration of participation and level of involvement in production) and (2) product characteristics (presence of audio-visual components such as still images, videos, music and narration, story theme, and length of the story).

Implementation Characteristics

We reviewed 8 implementation characteristics. We operationalized the implementation characteristics detailed by Peters et al [42] based on the framework by Proctor et al [43]. Implementation characteristics were acceptability, adoption, appropriateness, feasibility, fidelity, cost, coverage, and sustainability.

Report Characteristics

We included studies for which we could access the full-text reports, published in scholarly journals or unpublished in the case of dissertations and theses, written in English, and with no restrictions on country of origin or year of publication.

Search Methods

An exhaustive search was conducted in October 2019. We searched the following databases using a planned strategy to identify published studies: MEDLINE (Scopus), Embase (Scopus), PubMed, PsycINFO, Web of Science, CINAHL (EBSCO), Academic Search Complete (EBSCO), Abstracts in Social Gerontology (EBSCO), Psychology and Behavioral Sciences Collection (EBSCO), Health Source: Nursing Academic Edition (EBSCO), and SocINDEX (EBSCO). Unpublished studies were searched using ProQuest Dissertations and Theses and Open Access Theses and Dissertations. We also conducted backward citation tracking to search the reference lists of all the included studies to identify any relevant studies that may have been missed.

The selected search terms were chosen to describe the characteristics of the population and the activities necessary for the review. An example search (Scopus) is as follows:

> *TITLE-ABS-KEY (older adult* OR elder* OR older person* OR older people* OR dementia) AND TITLE-ABS-KEY (story OR stories OR storytelling OR biographi* OR biography*) AND TITLE-ABS-KEY (digital OR multimedia OR virtual).*

Data Collection and Analysis

Overview

Titles and abstracts produced by the database searches were collated using reference management software and duplicates were removed. Titles and abstracts were screened to remove obviously irrelevant reports before full texts of potentially relevant studies were assessed for inclusion based on the eligibility criteria. Using a pilot-tested data collection form, data were extracted from the included studies and synthesized. If there were multiple reports of a single study, they were identified and the extracted data were presented as findings from a single study. A total of 2 reviewers were involved in the screening of all the abstracts and full-text records and in the data extraction process. Discrepancies were resolved through discussion and consensus. Where necessary, a third reviewer was included in the discussion and a decision was made based on group consensus.

The corresponding authors of the studies were contacted via email for information to (1) clarify study eligibility for the review, (2) clarify or provide additional data to assist with data extraction, and (3) clarify or provide additional information to assist with quality assessment. If the authors could not be contacted to clarify study eligibility, the study was excluded. In instances where the authors could not be contacted for data extraction or quality assessment purposes, studies were included with missing data. Of the 8 authors, 5 (63%) authors responded to emails from the reviewers.

Owing to the considerable heterogeneity of study designs and study types, a statistical meta-analysis was not feasible. Data were synthesized using a qualitative content analysis guided by the framework provided by Popay et al [44]. Study findings were synthesized using textual descriptions and tabulation.

Qualitative studies were analyzed for themes by the first author (JS) [45].

Risk of Bias

The Mixed Methods Appraisal Tool (MMAT) was used to assess the methodological quality of the included studies [46]. The MMAT was chosen as it allowed for the appraisal of a variety of study designs, including quantitative nonrandomized, qualitative, and mixed methods studies. It comprised distinct sets of criteria to assess the validity of a study for each of the various study designs [47]. A total of 2 reviewers independently appraised all the studies and resolved the discrepancies through discussion. Where necessary, a third reviewer was included in the discussion and a decision was made based on group consensus. A critical discussion of the appraisal, both within and across studies, is presented.

Results

Study Identification

A PRISMA diagram of the selection process and flow of records at each stage is shown in Figure 1. Of the 391 records identified, duplicates were removed, and 248 (63.4%) records were screened for titles and abstracts. The full texts of 19.3% (48/248) of the records were reviewed for inclusion and exclusion criteria. Consensus was reached after independent review resulted in 90% agreement. Of the 10 records that met the eligibility criteria, 3 (30%) records were related to the same study. These 3 records were linked and presented as a single study.

Therefore, 8 studies were included in this review. Table 1 summarizes the study information, process characteristics, and product characteristics and Table 2 summarizes the key health-related outcomes of these studies. Table 3 presents the implementation characteristics.

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

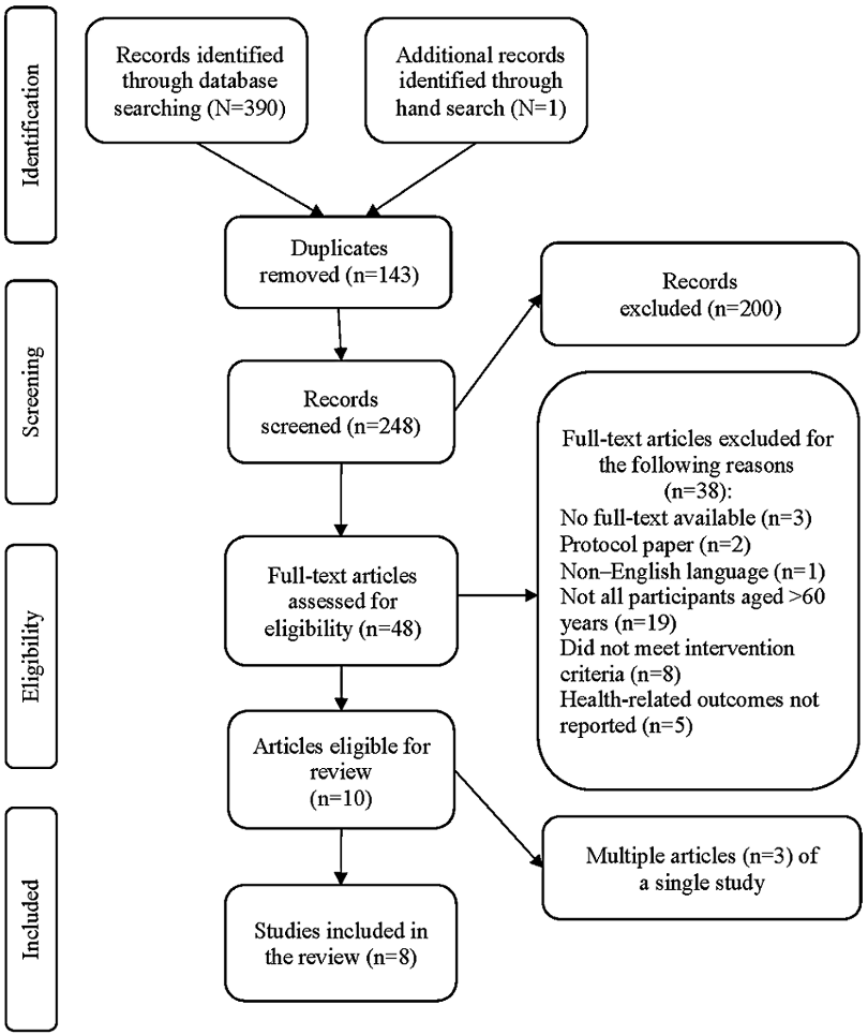


Table 1. Characteristics of the included studies (N=8).

Study	Study design	Sample description					Creation and content of stories
		Study population	Age (years), mean (range)	Diagnosis	Facility- or community-dwelling	Country	
Capstick et al [48]	Quantitative nonrandomized	10 (n=8, 80% female and n=2, 20% male)	87 (76-99)	Dementia	Long-term care facility	United Kingdom	Stories cocreated by participants and researchers; 1 hour per week for 6 weeks; stories consisted of still images (personal and generic), participant narration, music, and sound effects; 3.5-11 minutes in length.
Filoteo et al [49]	Quantitative nonrandomized	14	≥60	Dementia (mild to moderate)	Community-dwelling (recruited from an outpatient neuropsychological clinic)	United States	Stories created by family members, in a single day; stories consisted of still images (personal), family member voiceover, and music.
Subramaniam and Woods [33]	Mixed methods	6 (n=4, 67% female and n=2. 33% male)	82.2 (73-89)	Dementia (mild to moderate)	Long-term care facility	United Kingdom	Stories cocreated by participants and researchers; 1-1.5 hours per week over 7-10 weeks (mean 8.3 weeks); stories consisted of still and moving images (personal and generic), text captions, participant and family member voiceover, and music; 12-27 minutes in length (mean 18 minutes).
Crete-Nishihata et al [50], Damianakis et al [51], and Smith et al [32]	Qualitative	12 (n=7, 58% female and n=5, 42% male)	79.6 (60-95)	Alzheimer disease or MCI[a] (early to mid-stage)	Long-term care facility (n=2, 17%) and community-dwelling (n=10, 83%)	Canada	Stories cocreated by participants, family members, and 2 RAs[b]; 4-10 sessions over 2-12 months (mean 5.6 months); stories consisted of still and moving images (personal and generic), participant and RA voiceover, and music; 15-70 minutes in length (mean 39.1 minutes).
Critten and Kucirkova [52]	Qualitative	3 (n=1, 33% female and n=2, 67% male)	83.3 (72-94)	Dementia (mild to moderate)	Community-dwelling (recruited from a day center)	United Kingdom	Stories cocreated by participants and family members over 7 weeks; stories consisted of still images (personal and generic), text captions, and participant voiceover; user views story at their own pace—length varied.
O'Philbin [53]	Qualitative	6 (n=1, 17% female and n=5. 83% male)	Mean unknown (70-85)	Dementia (mild to moderate)	Community-dwelling	United Kingdom	Stories cocreated by participants and family members; 1-2 hours per week over 6 weeks.
Park et al [54]	Qualitative	7 (n=3, 43% female and n=4, 57% male)	74 (69-80)	Dementia (early stage)	Community-dwelling	Canada	Stories cocreated by participants and family members; 7 sessions of 2 hours over 6 weeks; stories consisted of still images (personal and generic), participant voiceover, and music; 3-8 minutes in length.
Sehrawat et al [55]	Qualitative	4 (n=3, 75% female and n=1, 25% male)	Mean unknown (73-82)	Healthy	Community-dwelling	United States	Stories cocreated by participants and young people; 6 sessions over 6 weeks, including a full-day workshop for production; stories consisted of still and moving images and participant voiceover.

[a]MCI: mild cognitive impairment.

[b]RA: research assistant.

Table 2. Key outcomes of included studies (N=8).

Study	Measures or tools used	Mood and affect	Memory	Quality of relationships	Social connectedness	Other health-related outcomes
Capstick et al [48]	BCC[a] coding frame (DCM[b]), BWP[c], and Arnstein Ladder of Citizen Participation	—[d]	—	—	Level of social citizenship increased by approximately 3 rungs.	Significant increase in positive well-being scores (P<.05) and significant decrease in negative indicators of well-being (P<.05) at midpoint. Well-being did not significantly decrease at 1 week after DS[e]; participants spent greater percentage of time engaged in reminiscence, conversation, and creative expression from before test to midpoint and after test.
Filoteo et al [49]	ET[f], STAI[g], HADS[h], NQOL[i], and CQ[j]	Statistically significant improvements on ET, STAI, HADS, and CQ from before test to after test (P<.05).	—	—	—	No statistically significant improvement on NQOL from before test to after test (P>.05).
Subramaniam and Woods [33]	QOL-AD[k], AMI[l], GDS[m], QCPR[n], and open-ended questionnaire	Improvement in scores on GDS at 4 weeks following the completion of DS.	Improvement in scores on AMI at 4 weeks following the completion of DS.	Improvement in scores on QCPR at 4 weeks following the completion of DS; participants, family members, and staff reported that the DS triggered memories and positive affect for the participant and enhanced interaction with family members and staff.	—	Improvement in scores on QOL-AD at 4 weeks following the completion of DS.
Crete-Nishihata et al [50], Damianakis et al [51], and Smith et al [32]	Semistructured interview and video recordings of screening sessions	Participants, family members, and staff reported emotional impacts of DS (eg, pleasure, sadness, and satisfaction); instances of positive emotion (n=291), negative emotion (n=6), and positive and negative emotion simultaneously (n=16).	Participants, family members, and staff reported that DS triggered long-term memories.	Participants, family members, and staff reported enhanced communication with family members and staff.	—	Participants, family members, and staff reported benefits for participants' sense of self.

Study	Measures or tools used	Mood and affect	Memory	Quality of relationships	Social connectedness	Other health-related outcomes
Critten and Kucirkova [52]	Interviews, field notes, and observations	Researchers reported the process was enjoyable for all participants and they experienced positive feelings of confidence, empowerment, and increased self-esteem.	—	—	—	—
O'Philbin [53]	Interviews	Participants and family members reported pride and enjoyment.	Participants and family members reported DS evoked memories.	—	—	—
Park et al [54]	Unstructured interviews, field notes, and audio recordings of sessions	Participants and family members reported enjoyment and a sense of accomplishment.	—	Researchers observed that participants were engaged in their relationships with their family members and the facilitator.	—	—
Sehrawat et al [55]	Open-ended questionnaire and unstructured interviews	—	—	—	Participants reported valued connections with young people and reported an increase in social connectedness and network size.	Participants found the process cathartic and therapeutic; however, they reported minimal to no change in physical and mental health.

[a]BCC: behavior category code.

[b]DCM: Dementia Care Mapping.

[c]BWP: Bradford Well-being Profile.

[d]Not addressed in the study.

[e]DS: digital story.

[f]ET: emotional thermometer.

[g]STAI: State-Trait Anxiety Inventory.

[h]HADS: Hospital Anxiety and Depression Scale.

[i]NQOL: Neuro–Quality of Life Depression Scale-modified.

[j]CQ: caregiver questionnaire.

[k]QOL-AD: Quality of Life in Alzheimer Disease scale.

[l]AMI: Autobiographical Memory Inventory.

[m]GDS: Geriatric Depression Scale.

[n]QCPR: Quality of the Caregiving Relationship Questionnaire.

Table 3. Implementation outcomes (N=8).

Study	Acceptability[a]	Adoption[b]	Appropriateness[c]	Feasibility[d]	Fidelity[e]	Cost[f]	Coverage[g]	Sustainability[h]
Capstick et al [48]	A participant became upset after watching her DS[i] that contained photos of a relative who had died.	—[j]	—	100% retention rate.	—	Use of free software (eg, Photo Story and Audacity).	—	The authors state that their step-by-step guide to participatory video is made available for others to replicate their work.
Filoteo et al [49]	—	—	—	—	—	"Low-cost" platform on a custom tablet for use on currently owned devices.	—	—
Subramaniam and Woods [33]	Reported no negative side effects. Enjoyed by all. Well received by relatives and staff. Disagreements with relatives regarding content and format were rarely encountered.	—	A total of 4 participants needed assistance to operate the DVD player; however, most reported preferring the digital form of their story over the previously made books. All participants needed someone to remind them to play the movie.	100% recruitment rate and 100% retention rate.	—	Used free software for production.	—	Widespread implementation requires consideration of time and skills—without the researcher, staff would have to take on the task of production.
Crete-Nishihata et al [50], Damianakis et al [51], and Smith et al [32]	There were varied viewing experiences, for example, after several viewings, 1 participant worried about how she could have made it differently and suggested that it should be editable. Strong rapport must be built among biographers, family members, and participants to resolve disagreements. Acknowledged that there are multiple interpretations of a life story and decisions must be made about including emotionally sensitive content. Personal media content is crucial.	—	Purposefully chose familiar technologies to enable easy adoption and integration (eg, television and DVD player). Still, some participants had trouble in operating the DVD player. Recognized that dementia severity may impact production participation.	52% recruitment rate (remaining participants declined owing to personal reasons) and 86% retention rate (1 dropout owing to death and 1 owing to time constraints).	—	Researchers worked for an average of 131.7 hours. As they became familiar with the process, they needed 60-90 hours to produce the DS. Family caregivers may not have time to do this without researchers. Production value varied, ranging from consumer-level to professional equipment. Inexpensive software was used.	—	A guide for families that may be interested in making their own DS is available.

Study	Acceptability[a]	Adoption[b]	Appropriateness[c]	Feasibility[d]	Fidelity[e]	Cost[f]	Coverage[g]	Sustainability[h]
Critten and Kucirkova [52]	Enjoyable for all participants who were all personally involved. The researcher– or carer–participant dynamic may influence the outcome of the study. Some may find this process difficult owing to sad memories.	—	Digital competence is necessary for participation as intended.	27% recruitment rate (remaining participants declined owing to technological limitations and time constraints) and 100% retention rate.	—	Used iPads and free software (Our Story).	—	—
O'Philbin [53]	For all participants, it was a mostly enjoyable experience. Some frustration was reported at not being able to recall specific things.	—	Some participants expressed "it's not for everyone" (life story work). The digital nature of the program was a barrier for some participants.	50% recruitment rate and 86% retention rate (1 dropout owing to declining to be interviewed).	—	—	—	The Book of You service does not check and encourage implementation with previous users, but the author suggests this could be considered.
Park et al [54]	Although participants were not able to explicitly address how they felt or specify what they enjoyed, there was a level of participation and enthusiasm that indicated interest.	—	The existing protocol for the workshop was modified in this study for people with dementia by having shorter and condensed sessions with a smaller participant group. A total of 2 participants did not have computers and were unable to use the program without assistance. None were able to use the program independently. A participant could not read her story aloud owing to visual impairment.	88% retention rate (1 dropout owing to time constraints).	—	Used freely available video software (WeVideo). Participants needed support from the facilitator and relatives to use the technology.	—	—

Study	Acceptability[a]	Adoption[b]	Appropriateness[c]	Feasibility[d]	Fidelity[e]	Cost[f]	Coverage[g]	Sustainability[h]
Sehrawat et al [55]	All participants enjoyed the process. Some negative emotions were produced—necessary to consider including a debrief in future and allowing more time for participants to share their stories in the group activity.	—	Workshop day was very long—participants began to tire and lose focus.	100% recruitment rate and 100% retention rate.	—	Funding was required to pay the student research assistants. Used free video software (WeVideo).	—	—

[a]Reported agreeableness or enjoyment by or on behalf of participants.

[b]The intention, initial decision, or action to try to use the activity.

[c]The perceived fit for the target group, reported by or on behalf of the target group.

[d]Reported rates of recruitment and retention.

[e]Whether the activity was implemented as it was designed to be.

[f]Costs associated with the implementation of activity (eg, financial costs, time, and human resource).

[g]The degree to which the population eligible to benefit from the activity actually receive it.

[h]Whether the activity was reported to be maintained in the given setting.

[i]DS: digital story.

[j]Not addressed in the study.

Study Characteristics

A total of 8 studies were reported in peer-reviewed journal articles (5/8, 63%), conference papers (2/8, 25%), and doctoral theses (1/8, 13%). Studies used qualitative (5/8, 63%), quantitative (2/8, 25%), and mixed methods (1/8, 13%) designs. The studies were based in the United Kingdom (4/8, 50%), the United States (2/8, 25%), and Canada (2/8, 25%). The quantitative and mixed methods studies used single-arm repeated measures (before-and-after) designs (3/8, 38%), of which 67% (2/3) of the studies used inferential statistics to test for statistical significance of pre–post differences in outcomes and 33% (1/3) of the studies showed descriptive statistics only. Qualitative studies used semistructured and unstructured interviews (6/8, 75%), of which 33% (2/6) also used field and observational notes. Qualitative analyses in these 6 studies were reported as content analysis (3/6, 50%) and thematic analysis (3/6, 50%). In 6 (75%) of the 8 studies, data on outcomes for participants and implementation characteristics were provided by participants and informants.

Quality assessment using the MMAT [46] indicated that study quality was acceptable overall and high for qualitative studies. Rating appraisal resulted in 75% agreement by 2 independent reviewers and consensus was reached via discussion. Studies were assessed on 5 quality criteria, with different sets of criteria for the qualitative, quantitative, and mixed methods studies. Of the 8 included studies, the 5 (63%) qualitative studies met all the MMAT criteria (5/5, 100%). For these studies, the qualitative approach was appropriate, using adequate data collection methods and presenting coherent findings that appeared to be adequately derived from and substantiated by data. The 25% (2/8) quantitative studies [48,49] did not account for confounders in their study design and analyses. In a third study, participants were identified as not representative of the target population. These studies were associated with MMAT scores of 3 and 4 out of 5. The mixed methods study did not explicitly produce an adequate rationale for using a mixed methods design and the quantitative component was assessed as not adhering to the quality criteria of the quantitative method, thus resulting in an MMAT score of 3 out of 5.

Participant Characteristics

The 8 studies comprised 62 participants, with study sample sizes ranging from 3 to 14 (mean 7.75, SD 3.88). Of the 88% (7/8) studies that provided information regarding participant gender, there were 27 female participants and 21 male participants. Age was reported inconsistently across the studies. Across 63% (5/8) studies that reported statistics on age (38/62, 63%), participants were aged between 60 and 99 years (mean 81 years). Cognitive status was not reported for participants in 13% (1/8) studies (4/62, 6%). Across the remaining 88% (7/8) studies, 52 participants were living with dementia, whereas 6 experienced MCI. In 63% (5/8) of the studies, dementia status was self-identified (32/62, 52%). Participants of 25% (2/8) of the studies (20/62, 32%) were assessed by the researchers for dementia status using the Diagnostic and Statistical Manual of Mental Disorders-IV or Diagnostic and Statistical Manual of Mental Disorders-V. Participants lived in long-term care facilities (18/62, 29%) or in the community (44/62, 71%). Community-dwelling participants were recruited from an outpatient neuropsychology clinic (14/44, 32%), local aging societies (11/44, 25%), day centers (9/44, 20%), or referred by health care professionals (2/44, 5%).

Methods of Digital Storytelling

Process Characteristics

Duration

The time taken to produce digital stories varied across studies. In 6 (75%) of the 8 studies, digital storytelling was conducted for 6-10 weeks. In 13% (1/8) of studies, stories were in production for up to 52 weeks [51]. In another study, digital stories were produced in 1 day [49].

Frequency

In 88% (7/8) studies, production sessions were generally held weekly. In 13% (1/8) of studies, participants produced their digital stories without structured sessions [52].

Length

Digital stories were produced in sessions lasting for 1-2 hours.

Producers

In 13% (1/8) of studies, the digital story was created by a family member without involving the older adult [49]. In 38% (3/8) studies, digital stories were cocreated by the older adults and the researchers. In 50% (4/8) studies, digital stories were cocreated by older adults, researchers, and family members.

Product Characteristics

Audio-Visual Composition

Across the 88% (7/8) studies that provided information regarding product composition, all digital stories included images that were both personal and generic, that is, stock photos sourced from the internet. In 86% (6/8) of these studies, a voiceover was also provided by participants. Of these 6 studies, 2 (33%) studies also included voiceover by family members and 1 (13%) included voiceover by research assistants. In 13% (1/8) of studies, stories were narrated by a family member only. Among the 8 included studies, music was included in 5 (63%) studies, moving images in 3 (38%) studies, and sound effects in 1 (13%) study.

Themes

Although all stories focused on participants' personal memories, 50% (4/8) of the studies identified specific themes of their participants' stories. Of the 8 studies, 1 (13%) study focused on significant places and events from the ages of 5-30 years [48] and another study (13%) focused on the areas of family and work [52]. In 25% (2/8) of the studies, digital stories explicitly took a broader focus and presented chronological accounts of life events, for example, childhood, teenage years, and career. [33,51].

Length

Stories differed in length; across the 50% (4/8) studies that indicated length, digital stories ranged from 3 minutes to 70 minutes.

Health-Related Outcomes

Health-related outcomes were explored across the included studies and clustered into five categories: mood and affect, memory, quality of relationships, social connectedness, and other health-related outcomes.

Mood and Affect

All 8 (100%) studies reported benefits related to mood and affect caused by digital storytelling activities.

Quantitative improvements in mood were reported in 25% (2/8) studies. Filoteo et al [49] reported that participants experienced significant (P<.05) pre–post improvements in anxiety, depression, overall emotional distress, and emotional functioning (as rated by family caregivers) after viewing their digital story. Subramaniam and Woods [33] found that participants reported a mean improvement in scores on standardized measures of depression (mean difference 1.84); however, such a difference was not evaluated for statistical significance.

Qualitatively, authors reported that digital storytelling fostered enjoyment [33,51-55] and other positive feelings including a sense of confidence [52], accomplishment, empowerment, self-esteem [52,54], enthusiasm [54], pleasure and satisfaction [51,55], and pride [53,54]. Some participants expressed grief and sorrow, but in the context of digital storytelling, this was considered in 13 (1/8) of studies as "natural expressions of loss, mitigated by the overall narrative of the life story" [33]. Damianakis et al [51] observed instances in which sadness and happiness were observed simultaneously.

Memory

Benefits associated with participant memory were reported across 88% (7/8) studies. Subramaniam and Woods [33] reported that participants experienced a mean improvement in scores on a standardized quantitative measure of autobiographical memory for factual knowledge (mean difference 8.92), providing some evidence for an effect of digital storytelling above and beyond the effect of the traditional life storybooks, which were created with participants before the digital storytelling activity. In contrast, autobiographical memory for specific events and incidents was overall highest following the traditional life storybook activity.

For participants across all 6 qualitative studies, digital storytelling provided a platform for stimulating long-term memories that may have been previously forgotten. Memories were elicited in various ways, including verbal prompts related to specific themes, for example, marriage and travel [51] or photographic material sourced from the internet [52]. Some participants and their family members identified that the digital story would serve as a valuable memory aid when their dementia progressed further [51,53].

Quality of Relationships

Across 63% (5/8) studies, digital storytelling activities improved participant relationships with their family members and professional caregivers.

The quality of caregiving relationships was assessed by Subramaniam and Woods [33], who reported mean improvements in scores on all subscales of a standardized quantitative measure (mean differences 0.83-6.83); however, such differences were not evaluated for statistical significance.

Qualitatively, 50% (4/8) studies described improved relationships between participants and family caregivers and

professional care staff during or after digital storytelling [33,51,53,54]. Family members interviewed by Damianakis et al [51] reported that digital storytelling facilitated enhanced communication with their relative living with dementia, in both the quality and quantity of their interactions, as past events were remembered and discussed. Professional carers who were interviewed in 13% (1/8) of studies acknowledged that viewing digital stories would help them better care for people living with dementia, as it enabled a deeper appreciation of their unique histories and provided relevant talking points [33].

Social Connectedness

Of the 8 included studies, 4 (50%) studies addressed the extent to which digital storytelling enhanced social connectedness. All the 4 (100%) studies reported that digital storytelling improved interactions among the participants, with others involved in cocreating stories or with viewers of the stories.

In 3 (75%) of these 4 qualitative studies, involvement in digital storytelling provided opportunities for increased social engagement [53-55]. Sehrawat et al [55] reported that older adults formed meaningful intergenerational connections with the students they were paired with over the 6-week activity by connecting through the shared experience of storytelling. Park et al [54] observed enhanced relationships between the participants and their family caregivers. Family caregivers of community-dwelling people living with dementia who attended a 6-week digital storytelling group workshop spoke of the social benefits associated with their relative meeting others [53].

In 2 (50%) of these 4 studies, broader social connections were examined. Participants involved in the intergenerational activity described sharing their digital stories beyond the activity with their friends and family, producing what the authors refer to as a *wave of connectedness* [55]. In providing evidence for participants' increased social citizenship, Capstick et al [48] reported that the digital stories of some participants were shared with the wider community (eg, on local history websites). Using a subjective measure of community engagement, the authors concluded that participants' potential for social citizenship improved owing to their engagement in digital storytelling. Furthermore, the authors presented a case study example of a participant who was taken out to the local theater to watch a play about cycling after the staff at the facility viewed her digital story, which was focused on her early passion for cycling to cope with her challenging experience growing up in a care facility.

Other Health Outcomes

The authors also reported that digital storytelling was associated with improvement in general well-being, quality of life, and sense of self and identity. Such stories were also seen to provide older adults with opportunities for leaving a legacy.

Quantitative improvements in well-being were reported by Capstick et al [48]. The authors reported a statistically significant improvement in well-being at the midpoint of the 6-week activity period ($P<.05$) and no significant decrease in well-being at 1 week following the activity ($P>.05$). On an observational measure, participants spent a greater percentage of time engaged in reminiscence, conversation, and creative expression at the midpoint compared with baseline. This pattern was maintained at 1 week following the end of the activity.

Filoteo et al [49] administered a standardized quantitative measure of quality of life and found no significant improvement following digital storytelling ($P>.05$). Subramaniam and Woods [29] reported a mean improvement in scores on a standardized quantitative measure of quality of life in Alzheimer disease; however, this was not tested for statistical significance.

A total of 5 studies proposed that the digital storytelling process served to elicit and validate a sense of self and identity, evident throughout production, for example, in selecting desired images and music to best represent their story [51], in shared viewing of their stories, [48], and simply in having their stories recorded in a tangible fashion [52]. Relatedly, people living with dementia and their family members also noted the value of the opportunity to leave a personal legacy [33,51,54].

Implementation Outcomes

Feasibility

Of the 8 included studies, 7 (88%) studies reported rates of recruitment or retention, indicating the potential for the feasibility of the digital storytelling activity. A small proportion of participants declined to be involved or dropped out of the studies owing to time constraints, death, difficulty in using the required technology, or other personal reasons (see Table 3 for the data).

Acceptability

All (8/8, 100%) studies reported that digital stories were agreeable and enjoyable. However, negative emotional reactions were noted in several studies, including some participants becoming upset during the activity, as the activity revived difficult memories and feelings of grief and loss [48,51,52,55]. Participants also became frustrated as they could not recall specific memories [53] and worried about how they could have made the story differently [51]. Notably, these instances of negative emotion were recorded as rare, occurring in only a small portion of participants per study.

Appropriateness

Of the 8 included studies, 6 (75%) studies discussed the appropriateness or the perceived fit of their activities. The primary consideration related to appropriateness was digital competence—authors noted that participants in some cases had difficulty in operating the required technology independently, affecting the production phase or their ability to view their digital story after the completion of the activity [33,51,52,54]. Of the 6 studies, 1 (17%) study cited that their protocol demanded too much attention and cognitive stamina of participants [55]. Some studies discussed the impact of dementia severity on the individual's capacity to participate as intended and noted the need to adopt a flexible approach [51,54].

Cost

Cost was typically referred to in the context of equipment needs. In most instances, the authors reported using inexpensive or freely available photo and video software to produce digital stories using devices already owned by researchers or

participants. Of the 8 studies, 1 (13%) study reported that funding was required to pay their student research assistants [55]. Only 13% (1/8) of studies provided an examination of the costs associated with time—researchers worked between 60 hours and 90 hours to produce lengthy digital stories, ranging from consumer-level to professional quality—with the authors acknowledging that family members may not have the time to undertake this activity themselves [51].

Fidelity

Studies did not assess whether the activity implement was as intended; fidelity of the protocols was not measured. Of the 8 included studies, 1 (13%) study stated that the authors' existing protocol for the digital storytelling workshop was modified for the current sample (people living with dementia) before commencement of the activity [54]; however, even in this study, fidelity of the modified protocol was not assessed.

Adoption, Coverage, and Sustainability

The authors did not comment on adopting the activity in routine practice at the individual or organizational level or the coverage of the activity. Similarly, the authors did not report whether digital storytelling activities were sustained or maintained in their respective settings. However, the authors of 25% (2/8) of studies reported that they had prepared digital storytelling guidebooks that were available upon request for those interested in adopting their approaches [48,51].

Discussion

Principal Findings

The purpose of this systematic review was to explore the characteristics, outcomes, and potential for implementation of digital storytelling with older adults. This review summarizes the methods used to produce digital stories, features of the digital story products, health-related outcomes of digital storytelling, and implementational considerations of digital storytelling activities. The 8 studies that were reviewed comprised 62 participants aged between 60 and 99 years, most of whom had a diagnosis of dementia or MCI and lived in the community.

The review adopted an overly inclusive definition of digital storytelling, whereby restrictions were not placed on the level of participant involvement in creating stories, time taken to produce stories, or the length of the stories themselves, provided they satisfied all other inclusion criteria. Studies used markedly different methods for producing digital stories; in most studies, digital stories were cocreated by older adults and researchers and family members over several weeks. Some were short (eg, 3-5 minutes), consistent with Lambert's protocols [21] and others were significantly longer, such as the *multimedia biographies* that were up to 70 minutes in length [32]. Digital stories were centered on the lived experiences of older adults, including stories of family, work, travel, and significant places and events. All digital stories used personal and generic images and voiceover narration, with many including music and some including moving images or video and sound effects.

Digital storytelling was associated with four overarching health-related outcomes: positive mood and affect; improved memory; enhanced relationships among older adults, family, and professional caregivers; and improved social connectedness. Approaches to assessing outcomes were heterogeneous, with outcomes assessed using a variety of quantitative and qualitative methods.

The reported potential for implementation varied across the studies. The included studies did not explicitly aim to examine the potential for the implementation of their activities. Using the implementation framework presented by Peters et al [42], the review found some evidence for the acceptability of digital storytelling through overall participant agreeableness and enjoyment. However, some authors reported noteworthy issues, including unanticipated negative emotions in few participants. Studies delivered activities that were considered appropriate for the target population; however, some issues related to fit were reported, including participants' poor digital literacy and cognitive and emotional demands. There was evidence for feasibility as retention was relatively high and recruitment rates were adequate. Activities were delivered with a low financial cost; however, in some instances, the time commitment required for researchers, and research assistants, and family members was considerable. Studies did not discuss the adoption of the activity at the individual or organizational level, sustainability of the activity, or its coverage. Overall, findings from this review suggest that digital storytelling is implementable when activities are designed with careful consideration of the physical, cognitive, and emotional needs of the target population.

Limitations

The MMAT [46] demonstrated that the included studies were generally of high quality.

However, several important questions remain unanswered. The single-group repeated measure designs of the quantitative studies pose a low level of evidence [56]. As none of the included studies used a comparator group, conclusions regarding the efficacy of digital storytelling compared with other or no activity cannot be made with confidence. Notably, the studies also did not conduct follow-up assessments to explore whether the effects were sustained for greater than a week [47] following digital storytelling. Studies did not generally aim to explore or account for confounding factors in their analyses—it remains largely unknown what components of the digital storytelling process, such as social interaction, stimulating and sharing of memories, feeling heard and valued, and producing a tangible digital story, have the most effect on outcomes. In addition, it remains unknown whether the outcomes of the digital storytelling process (after creation) are distinct from those related to digital story viewing (after viewing). Although some studies assessed outcomes at several time points, this question was not explicitly addressed and remains to be explored further. Nearly all participants (58/62, 93%) in the pool of reviewed studies were living with dementia or MCI. Hence, findings from this review may not be generalizable to cognitively healthy older adults.

Owing to the considerable heterogeneity of the studies, a qualitative content analysis was conducted to synthesize the evidence presented. The overly inclusive definition of digital storytelling was necessary to capture all relevant studies in this emerging literature; however, the robustness of the synthesis is

moderately limited by the markedly differing purposes of digital storytelling and heterogeneous outcome measurement across only a few studies. A larger number of homogenous studies would allow a more confident account of the outcomes of digital storytelling.

Although an exhaustive search method was used, a librarian with expertise in search strategies was not consulted and forward citation tracking was not conducted. Gray literature, besides unpublished theses and conference papers, was not searched.

Furthermore, the primary focus of this review and the search criteria was to explore the health-related outcomes of digital storytelling activities—studies were only included if they reported at least one health-related outcome. Hence, studies were not included in this review if they did not report health outcomes; studies that only reported on methods or implementation potential of digital storytelling activities were not included in the current review.

Conclusions

This is the first review to systematically survey the current state of digital storytelling literature for older adults. Despite varied approaches, the review found that, when used with older adults, digital storytelling is largely acceptable and feasible and shows potential for benefits related to mood and affect, memory, quality of relationships, and social engagement.

Acknowledgments

The authors wish to thank Kathleen de Boer for her assistance with data extraction. This research was funded by the Swinburne University Postgraduate Research Award.

Conflicts of Interest

None declared.

References

1. United Nations, Department of Economic and Social Affairs, Population Division. World Population Ageing 2017 - Highlights (ST/ESA/SER.A/397). New York: United Nations; 2017.
2. Braam AW, Copeland JR, Delespaul PA, Beekman AT, Como A, Dewey M, et al. Depression, subthreshold depression and comorbid anxiety symptoms in older Europeans: results from the EURODEP concerted action. J Affect Disord 2014 Feb;155:266-272. [doi: 10.1016/j.jad.2013.11.011] [Medline: 24355647]
3. Nicholson NR. A review of social isolation: an important but underassessed condition in older adults. J Prim Prev 2012 Jun;33(2-3):137-152. [doi: 10.1007/s10935-012-0271-2] [Medline: 22766606]
4. Perissinotto CM, Stijacic CI, Covinsky KE. Loneliness in older persons: a predictor of functional decline and death. Arch Intern Med 2012 Jul 23;172(14):1078-1083 [FREE Full text] [doi: 10.1001/archinternmed.2012.1993] [Medline: 22710744]
5. Paddock K, Wilson C, Walshe C, Todd C. Care home life and identity: a qualitative case study. Gerontologist 2019 Jul 16;59(4):655-664 [FREE Full text] [doi: 10.1093/geront/gny090] [Medline: 30085052]
6. Lee VS, Simpson J, Froggatt K. A narrative exploration of older people's transitions into residential care. Aging Ment Health 2013;17(1):48-56. [doi: 10.1080/13607863.2012.715139] [Medline: 22913579]
7. Bhar S. Reminiscence therapy: a review. In: Pachana NA, Laidlaw K, editors. The Oxford Handbook of Clinical Geropsychology. New York: Oxford University Press; 2014.
8. Webster JD, Bohlmeijer ET, Westerhof GJ. Mapping the future of reminiscence: a conceptual guide for research and practice. Res Aging 2010 May 24;32(4):527-564. [doi: 10.1177/0164027510364122]
9. East L, Jackson D, O'Brien L, Peters K. Storytelling: an approach that can help to develop resilience. Nurse Res 2010;17(3):17-25. [doi: 10.7748/nr2010.04.17.3.17.c7742] [Medline: 20450085]
10. Mager B. Storytelling contributes to resilience in older adults. Activit Adapt Aging 2018 Mar 12;43(1):23-36. [doi: 10.1080/01924788.2018.1448669]
11. Menn L, Corsten S, Lauer N, Wallace SJ. The effectiveness of biographical approaches in long-term care: a systematic review. Gerontologist 2020 May 15;60(4):309-328. [doi: 10.1093/geront/gnz074] [Medline: 31175820]
12. Shropshire M. Reminiscence intervention for community-dwelling older adults without dementia: a literature review. Br J Community Nurs 2020 Jan 02;25(1):40-44. [doi: 10.12968/bjcn.2020.25.1.40] [Medline: 31874077]
13. Bohlmeijer E, Roemer M, Cuijpers P, Smit F. The effects of reminiscence on psychological well-being in older adults: a meta-analysis. Aging Ment Health 2007 May;11(3):291-300. [doi: 10.1080/13607860600963547] [Medline: 17558580]
14. Pinquart M, Forstmeier S. Effects of reminiscence interventions on psychosocial outcomes: a meta-analysis. Aging Ment Health 2012;16(5):541-558. [doi: 10.1080/13607863.2011.651434] [Medline: 22304736]
15. Redulla R. Reminiscence therapy for dementia. Issues Ment Health Nurs 2020 Mar;41(3):265-266. [doi: 10.1080/01612840.2019.1654572] [Medline: 31714840]
16. Cuevas PE, Davidson PM, Mejilla JL, Rodney TW. Reminiscence therapy for older adults with Alzheimer's disease: a literature review. Int J Ment Health Nurs 2020 Jun;29(3):364-371. [doi: 10.1111/inm.12692] [Medline: 31984570]
17. Thompson R. Using life story work to enhance care. Nurs Older People 2011 Oct;23(8):16-21. [doi: 10.7748/nop2011.10.23.8.16.c8713] [Medline: 22017158]

18. Doran C, Noonan M, Doody O. Life-story work in long-term care facilities for older people: an integrative review. J Clin Nurs 2019 Apr;28(7-8):1070-1084. [doi: 10.1111/jocn.14718] [Medline: 30431682]
19. Cooney A, O'Shea E. The impact of life story work on person-centred care for people with dementia living in long-stay care settings in Ireland. Dementia (London) 2019;18(7-8):2731-2746. [doi: 10.1177/1471301218756123] [Medline: 29411663]
20. Grøndahl VA, Persenius M, Bååth C, Helgesen AK. The use of life stories and its influence on persons with dementia, their relatives and staff - a systematic mixed studies review. BMC Nurs 2017;16:28 [FREE Full text] [doi: 10.1186/s12912-017-0223-5] [Medline: 28588424]
21. Lambert J. Digital Storytelling: Capturing Lives, Creating Community. Oxfordshire, United Kingdom: Routledge; 2010.
22. Robin BR. Digital storytelling: a powerful technology tool for the 21st century classroom. Theory Pract 2008 Jul 11;47(3):220-228. [doi: 10.1080/00405840802153916]
23. Sadik A. Digital storytelling: a meaningful technology-integrated approach for engaged student learning. Edu Technol Res Dev 2008 Apr 11;56(4):487-506. [doi: 10.1007/s11423-008-9091-8]
24. Lenette C. Using Digital Storytelling in Participatory Research With Refugee Women. London, UK: SAGE Publications Ltd; 2017.
25. Gubrium A. Digital storytelling: an emergent method for health promotion research and practice. Health Promot Pract 2009 Apr;10(2):186-191. [doi: 10.1177/1524839909332600] [Medline: 19372280]
26. Rose CB. The subjective spaces of social engagement: cultivating creative living through community-based digital storytelling. Psychoanal Cult Soc 2016 Dec 21;21(4):386-402. [doi: 10.1057/pcs.2015.56]
27. Laing CM, Moules NJ, Estefan A, Lang M. "Stories Take Your Role Away From You": understanding the impact on health care professionals of viewing digital stories of pediatric and adolescent/young adult oncology patients. J Pediatr Oncol Nurs 2017;34(4):261-271. [doi: 10.1177/1043454217697023] [Medline: 28376686]
28. Willox AC, Harper SL, Edge VL. Storytelling in a digital age: digital storytelling as an emerging narrative method for preserving and promoting indigenous oral wisdom. Qual Res 2012 Oct 22;13(2):127-147. [doi: 10.1177/1468794112446105]
29. Gubrium A, Fiddian-Green A, Lowe S, DiFulvio G, Peterson J. Digital storytelling as critical narrative intervention with adolescent women of Puerto Rican descent. Crit Public Health 2019;29(3):290-301. [doi: 10.1080/09581596.2018.1451622] [Medline: 31130780]
30. Lambert J. Where it all started: the center for digital storytelling in California. In: Story Circle: Digital Storytelling around the World. Hoboken, New Jersey, United States: Blackwell Publishing; 2009:77-90.
31. Fiddian-Green A, Kim S, Gubrium AC, Larkey LK, Peterson JC. Restor(y)ing health: a conceptual model of the effects of digital storytelling. Health Promot Pract 2019 Jul;20(4):502-512. [doi: 10.1177/1524839918825130] [Medline: 30736703]
32. Smith KL, Crete-Nishihata M, Damianakis T, Baecker RM, Marziali E. Multimedia biographies: a reminiscence and social stimulus tool for persons with cognitive impairment. J Technol Hum Serv 2009 Nov 09;27(4):287-306 [FREE Full text] [doi: 10.1080/15228830903329831]
33. Subramaniam P, Woods B. Digital life storybooks for people with dementia living in care homes: an evaluation. Clin Interv Aging 2016;11:1263-1276 [FREE Full text] [doi: 10.2147/CIA.S111097] [Medline: 27698556]
34. Stenhouse R, Tait J, Hardy P, Sumner T. Dangling conversations: reflections on the process of creating digital stories during a workshop with people with early-stage dementia. J Psychiatr Ment Health Nurs 2013 Mar;20(2):134-141. [doi: 10.1111/j.1365-2850.2012.01900.x] [Medline: 22413774]
35. Hausknecht S, Vanchu-Orosco M, Kaufman D. Digitising the wisdom of our elders: connectedness through digital storytelling. Ageing Society 2018 Jul 17;39(12):2714-2734. [doi: 10.1017/S0144686X18000739] [Medline: 30886898]
36. Hewson J, Danbrook C, Sieppert J. Engaging post-secondary students and older adults in an intergenerational digital storytelling course. Contemp Issues Educ Res 2015 Jul 13;8(3):135-142. [doi: 10.19030/cier.v8i3.9345]
37. Loe M. The digital life history project: intergenerational collaborative research. Gerontol Geriatr Educ 2013;34(1):26-42. [doi: 10.1080/02701960.2012.718013] [Medline: 23362853]
38. Matheson-Monnet C. Reminiscence therapy and intergenerational interventions for enhancing self-identity and social inclusion of older people and people living with dementia (PLDs). In: Papa R, editor. Handbook on Promoting Social Justice in Education. Cham: Springer; 2020:687-710.
39. Stargatt J, Bhar S, Bhowmik J, Al Mahmud A. Implementing digital storytelling for health-related outcomes in older adults: protocol for a systematic review. JMIR Res Protoc 2019 Dec 20;8(12):e15512 [FREE Full text] [doi: 10.2196/15512] [Medline: 31859687]
40. Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst Rev 2015 Jan;4:1 [FREE Full text] [doi: 10.1186/2046-4053-4-1] [Medline: 25554246]
41. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Br Med J 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
42. Peters DH, Adam T, Alonge O, Agyepong IA, Tran N. Implementation research: what it is and how to do it. Br Med J 2013 Nov 20;347:f6753. [doi: 10.1136/bmj.f6753] [Medline: 24259324]

43. Proctor E, Silmere H, Raghavan R, Hovmand P, Aarons G, Bunger A, et al. Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. Adm Policy Ment Health 2011 Mar;38(2):65-76 [FREE Full text] [doi: 10.1007/s10488-010-0319-7] [Medline: 20957426]
44. Popay J, Roberts H, Sowden A, Petticrew M, Arai L, Rodgers M, et al. Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC Methods Programme. Lancaster University 2006. [doi: 10.13140/2.1.1018.4643]
45. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
46. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Edu Inform 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
47. Hong QN, Pluye P, Fàbregues S, Bartlett G, Boardman F, Cargo M, et al. Improving the content validity of the mixed methods appraisal tool: a modified e-Delphi study. J Clin Epidemiol 2019 Jul;111:49-59 [FREE Full text] [doi: 10.1016/j.jclinepi.2019.03.008] [Medline: 30905698]
48. Capstick A, Ludwin K, Chatwin J, Walters E. Participatory video and well-being in long-term care. J Dement Care 2016;24(1):26-29 [FREE Full text]
49. Filoteo JV, Cox EM, Split M, Gross M, Culjat M, Keene D. Evaluation of reminx as a behavioral intervention for mild to moderate dementia. Annu Int Conf IEEE Eng Med Biol Soc 2018 Jul;2018:3314-3317. [doi: 10.1109/EMBC.2018.8513049] [Medline: 30441098]
50. Crete-Nishihata M, Baecker RM, Massimi M, Ptak D, Campigotto R, Kaufman LD, et al. Reconstructing the past: personal memory technologies are not just personal and not just for memory. Hum-Comput Interact 2012;27(1-2):92-123 [FREE Full text]
51. Damianakis T, Crete-Nishihata M, Smith KL, Baecker RM, Marziali E. The psychosocial impacts of multimedia biographies on persons with cognitive impairments. Gerontologist 2010 Feb;50(1):23-35. [doi: 10.1093/geront/gnp104] [Medline: 19592639]
52. Critten V, Kucirkova N. 'It Brings it all Back, all those Good Times; it Makes Me Go Close to Tears'. Creating digital personalised stories with people who have dementia. Dementia (London) 2019 Apr;18(3):864-881. [doi: 10.1177/1471301217691162] [Medline: 28161989]
53. Philbin LO. Reminiscence and digital life story work for dementia care. PhD Thesis, Bangor University. 2019. URL: https://tinyurl.com/2p9hkjs6 [accessed 2021-12-28]
54. Park E, Owens H, Kaufman D, Liu L. Digital storytelling and dementia. In: Proceedings of the International Conference on Human Aspects of IT for the Aged Population. 2017 Presented at: International Conference on Human Aspects of IT for the Aged Population; July 9-14, 2017; Vancouver, Canada p. 443-451. [doi: 10.1007/978-3-319-58536-9_35]
55. Sehrawat S, Jones C, Orlando J, Bowers T, Rubins A. Digital storytelling: a tool for social connectedness. Gerontechnology 2017 Apr 28;16(1):56-61. [doi: 10.4017/gt.2017.16.1.006.00]
56. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. Plast Reconstr Surg 2011 Jul;128(1):305-310 [FREE Full text] [doi: 10.1097/PRS.0b013e318219c171] [Medline: 21701348]

Abbreviations

MCI: mild cognitive impairment
MMAT: Mixed Methods Appraisal Tool
PRISMA: Preferred Reporting Items for Systematic Review and Meta-Analyses
PRISMA-P: Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols

Edited by R Kukafka; submitted 22.02.21; peer-reviewed by P Pluye, A Hashim, B Hattink; comments to author 14.04.21; revised version received 28.04.21; accepted 22.10.21; published 12.01.22.

Please cite as:
Stargatt J, Bhar S, Bhowmik J, Al Mahmud A
Digital Storytelling for Health-Related Outcomes in Older Adults: Systematic Review
J Med Internet Res 2022;24(1):e28113
URL: https://www.jmir.org/2022/1/e28113
doi:10.2196/28113
PMID:35019845

Review

Direct Access for Patients to Diagnostic Testing and Results Using eHealth: Systematic Review on eHealth and Diagnostics

Anke Versluis[1,2], PhD; Kyma Schnoor[1,2,3], MSc; Niels H Chavannes[1,2], MD, PhD; Esther PWA Talboom-Kamp[1,2,3], MBA, MD, PhD

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

[2]National eHealth Living Lab, Leiden University Medical Center, Leiden, Netherlands

[3]Saltro Diagnostic Center, Utrecht, Netherlands

Corresponding Author:
Kyma Schnoor, MSc
Department of Public Health and Primary Care
Leiden University Medical Center
Albinusdreef 2
Leiden, 2300 RC
Netherlands
Phone: 31 71526 8433
Email: k.schnoor@lumc.nl

Abstract

Background: The number of people with chronic diseases and the subsequent pressure on health care is increasing. eHealth technology for diagnostic testing can contribute to more efficient health care and lower workload.

Objective: This systematic review examines the available methods for direct web-based access for patients to diagnostic testing and results in the absence of a health care professional in primary care.

Methods: We searched the PubMed, Embase, Web of Sciences, Cochrane Library, Emcare, and Academic Search Premier databases in August 2019 and updated in July 2021. The included studies focused on direct patient access to web-based triage leading to diagnostic testing, self-sampling or testing, or web-based communication of test results. A total of 45 studies were included. The quality was assessed using the Mixed Methods Appraisal Tool.

Results: Most studies had a quantitative descriptive design and discussed a combination of services. Diagnostic test services mainly focused on sexually transmitted infections. Overall, the use was high for web-based triage (3046/5000, >50%, who used a triage booked a test), for self-sampling or self-testing kits (83%), and the result service (85%). The acceptability of the test services was high, with 81% preferring home-based testing over clinic-based testing. There was a high rate of follow-up testing or treatment after a positive test (93%).

Conclusions: The results show that direct access to testing and result services had high use rates, was positively evaluated, and led to high rates of follow-up treatment. More research on cost-effectiveness is needed to determine the potential for other diseases. Direct access to diagnostic testing can lower the threshold for testing in users, potentially increase efficiency, and lower the workload in primary care.

(J Med Internet Res 2022;24(1):e29303) doi:10.2196/29303

KEYWORDS

eHealth; systematic review; diagnostic testing; home-based test; self-test

Introduction

Background

As the population ages and the number of people with chronic diseases increase, the pressure on the health care system continues to rise [1,2]. This increased pressure is particularly noticeable in primary care where, over the years, the workload had already increased because of health care transformations. Primary care physicians, for example, are required to perform more preventive and complex care, work more according to evidence-based guidelines, and focus on person-centered care delivery [3,4]. Thus, physicians are required to do more in less time, and this increased workload can negatively affect the quality of patient care [4,5] and result in lower levels of job

satisfaction of health care professionals (HCPs) [6,7]. Care delivery needs to be reformed to meet the needs of an aging population.

eHealth has been identified as a potential method to make health care delivery more efficient and can thereby help to decrease the workload [8,9]. eHealth can be defined as "health services and information delivered or enhanced through the Internet and related technologies" [10,11]. Currently, different eHealth applications are used to different extents in primary care. The advantage of eHealth applications is that health care delivery can be more efficient and can operate partially, or even completely, independent of the HCP. Gaining more insight into how eHealth is used in primary care can help to identify promising approaches that may help to lower the workload in primary care and contribute to better health care quality.

Requesting laboratory diagnostic testing, which refers to testing to determine the presence of a disease, and the communication of the results has shown promise for digitization. Indeed, eHealth technology has been applied successfully in the three stages of laboratory diagnostic testing. The first stage is *triage and advice on diagnostic testing*, where typically an HCP asks the patient a set of questions to determine whether and what diagnostic tests are relevant. An example of web-based triage was provided by Polilli et al [12], who used a web-based questionnaire (ie, triage) to determine an individual's risk for HIV and sexually transmitted infections (STIs). On the basis of the calculated risk, individuals were automatically linked to nearby testing and counseling facilities. The second stage is the actual *testing* (eg, a blood test is performed to determine the presence of an infection). There have now been initiatives where laboratory tests can be ordered on the internet and are shipped to the individual for self-testing or self-sampling [13,14]. Self-testing refers to an approach in which individuals can collect their specimen (eg, blood) and interpret the results using a rapid diagnostic test. In self-sampling, individuals collect their specimens, but the specimen is tested elsewhere (eg, laboratory). The third stage is the *communication of test results* to the patient. A course of action is then determined based on the results. Instead of having the HCP communicate the results, it can also be communicated on the web or via an app, independent of the professional. Automated SMS text messages can be used to deliver tuberculosis testing results [15] or negative HIV test results can be automatically reported using the internet or a voicemail system. To our knowledge, a comprehensive overview of the different methods used to provide patients with direct web-based access to laboratory diagnostic testing and results is not yet available.

Objective

The aim is to conduct a systematic review to identify and summarize the available methods for direct web-based access for participants to diagnostic testing and results in the absence of an HCP in primary care. The available reviews show promise (eg, suggesting that self-tests are acceptable and can increase the uptake and frequency of testing) [16,17], but are limited to self-sampling and self-testing and do not include other forms of digitization. Moreover, the existing reviews focus on specific populations such as men who have sex with men (MSM) [18,19] or on specific health conditions such as HIV or chlamydia [20,21]. To widen the scope, this systematic review will include studies focusing on digitization in one or more phases of laboratory diagnostic testing. Specifically, studies that focus on direct access for patients to (1) web-based triage that leads to diagnostic testing, (2) self-sampling or testing, or (3) the test results are included (or both). The review was not restricted to specific populations or health conditions. Identification and summary of possible methods for direct access to diagnostic testing and result services will help identify usable and effective methods that can potentially increase the accessibility and cost-effectiveness of health care and simultaneously reduce the workload of primary care professionals.

Methods

The PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines for reporting systematic reviews were used [22]. The systematic review was not registered, but a strict protocol was used to search and select studies and to select data.

Search Strategies

PubMed, Embase, Web of Science, Cochrane Library, Emcare, and Academic Search Premier were searched on August 16, 2019, to identify publications about digitization in the laboratory diagnostic setting (ie, web-based triage that leads to laboratory testing, self-sampling or testing, or web-based communication of laboratory test results). The search was updated on July 21, 2021. Search terms related to laboratory diagnostics and eHealth were combined (see Multimedia Appendix 1 for the full search strings). The search was limited to peer-reviewed publications. The reference lists of relevant reviews and the selected publications were also searched.

Study Selection

The titles and abstracts of the identified publications were screened for relevance. The full text was screened when it concerned potentially relevant publications or when there was insufficient information in the abstract to adequately assess the relevance. Several inclusion criteria were used to select the relevant publications. First, the publication should focus on a *specific* web-based laboratory diagnostic service. The service could be (1) a web-based questionnaire or triage that directs users to a laboratory test (in the clinic or at home), (2) an ordered self-sampling or testing kit, or (3) a system for web-based communication of laboratory test results to users. Second, the laboratory diagnostic service should be (partly) independent of an HCP (eg, the questionnaire or triage should not be administered over the phone by the HCP; the test kit should not be provided in-person; administering the test should not require assistance from an HCP; and the test results should not be communicated through a phone call). Regarding the latter, the publication was included when it discussed a result service that was partly independent of an HCP (ie, negative test results were automatically communicated and, in case of positive test results, there was contact between the HCP and patient). Third, the publication should focus on primary care settings; however, this exclusion criterion was omitted for studies conducted in Africa (as there is no clear distinction between primary and

secondary care). Fourth, the study outcomes should specifically examine the laboratory diagnostic service (ie, the triage, test, or web-based communication of the test results) and not the surrounding procedures (eg, the acceptability of the consent procedure or the development of the service). Relevant outcomes included actual use or uptake, feasibility and acceptability, and effectiveness (eg, the time taken to test for diagnosis, understanding of test results, and the accuracy of triage). Publications were excluded if the laboratory diagnostic service focused on (national) screening campaigns, the monitoring of disease progression, or retesting or increasing retesting rates. Reviews, trial protocols, non–peer-reviewed papers, non-English papers, and publications without data or with only hypothetical data were also excluded. AV screened all the titles, and AV and ET independently screened the abstracts and full-text publications. For the second search, which was used to update the data, KS screened all the titles. The screening of abstracts was performed independently by AV and KS, and full-text publication screening was performed independently by KS and ET. Discrepancies were resolved through discussion.

Coding

A standardized coding form was used to extract all relevant information from the identified publications. The following information was extracted: (1) the first author and publication year, (2) the country in which the study was conducted, (3) the type of study design (using the classification by Hong et al [23]), and (4) sample characteristics (ie, target group, sample size, age, and gender). It was then determined which laboratory diagnostic service was studied (ie, web-based triage, self-sampling or testing, web-based result service, or any combination of the former three options). The names of the web-based laboratory diagnostic service and the recruitment method were also coded. The different recruitment methods were categorized as social marketing (eg, media, social media, magazines, flyers, advertisements, or promotion in target groups), community outreach (eg, face-to-face recruitment and community events), health service recruitment (ie, direct recruitment by the service provider in past service users), and other recruitment strategies. Details of the laboratory diagnostic services were extracted. Different data were collected based on what services or combinations of services were studied. For the web-based triage service, the aim of the triage was extracted, and it was determined whether it resulted in clinic- or home-based testing (ie, self-sampling or self-testing). For the self-sampling or self-testing service, the following information was extracted when applicable: (1) type of test (ie, self-sampling or self-testing); (2) for what disease; (3) type of specimen (eg, urine specimen); (4) method of how the test kit was ordered, delivered, and how the specimen could be returned; (5) method of instruction (ie, written or video); and (6) costs. For the web-based result service, we coded the method of result notification (eg, on the web or email), whether the notification was entirely or partially independent from an HCP, the average number of days before results were communicated, and whether individuals with positive results were linked to follow-up confirmatory testing or treatment. Results were then extracted, specifically results related to the service evaluation (see the *Study Selection* section) and not, for example, the characteristics of the service users. AV carried out the coding, and ET independently coded a subsample. There was substantial agreement between the 2 authors (ie, 77%). For the second search, the update, coding was done by KS.

Quality Assessment

The quality of the included studies was assessed using the valid Mixed Method Appraisal Tool (MMAT) [23]. This tool was able to assess the quality of different study designs. The MMAT was chosen because it can be used to assess the methodological quality of 5 different study designs, specifically qualitative, randomized controlled, nonrandomized, quantitative descriptive, and mixed methods studies. The design was determined for each publication, and 5 corresponding quality criteria were rated. The criteria are shown in Multimedia Appendix 2. Each item was rated with *yes* (ie, indicative of good quality), *no* (ie, indicative of poor quality), or *can't tell* (ie, insufficient evidence to determine the quality).

Furthermore, a numeric score was calculated to provide insight into the overall quality of each study. The AV conducted the complete quality assessment, and ET assessed a 10% subsample. The average Cohen κ was 0.80, indicating strong interrater reliability [24]. For the second search, KS completed the quality assessment of the studies (n=6).

Data Analysis

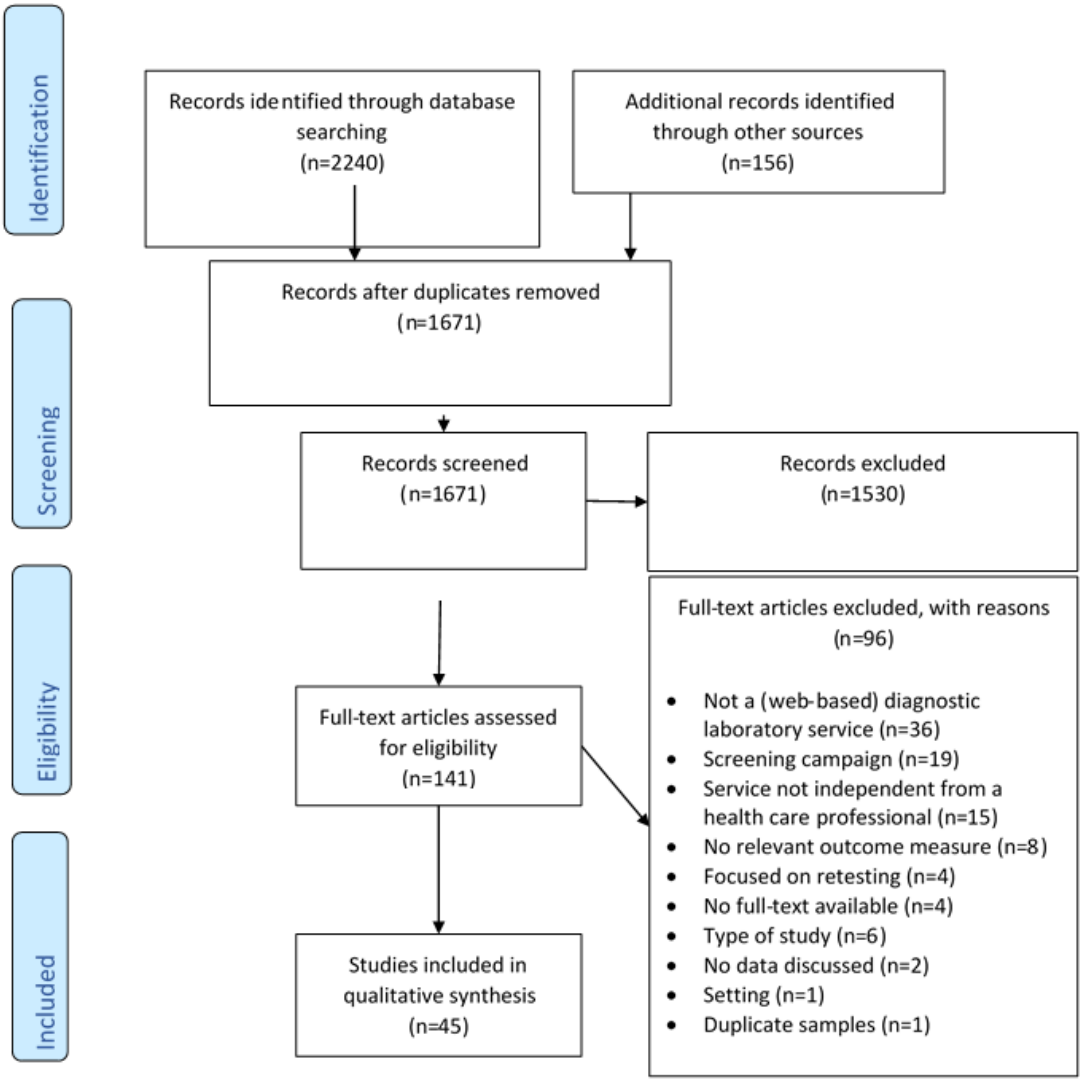
Data were extracted from the results sections of the studies, as described in the coding paragraph. Relevant outcome measures were extracted verbatim and added to the database, enabling the clustering of different outcome measures. The main findings are presented separately for the different service types. A detailed description of the findings of the included studies is provided in Multimedia Appendix 3 [12-15,25-65].

Results

Study Selection

As shown in Figure 1, the 2 search strategies resulted in 1671 publications after removing duplicates. The titles and abstracts were screened for relevance, and the full texts of 141 publications were checked. A total of 96 publications were excluded, most frequently, because the publication did not report on a (web-based) diagnostic laboratory service (n=36), it concerned a national screening campaign (n=19), or the service was not independent of an HCP (n=15). Finally, 45 publications were included in the qualitative synthesis, and 6 studies were included in the second search.

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



Study Characteristics

Most of the included studies had a quantitative descriptive design (n=28) [12,13,15,25-50]. In the remaining studies, a (quantitative) nonrandomized design was reported 6 times [32,51-55], a randomized controlled design was reported 5 times [56-60], a mixed methods design was reported 3 times [14,61,62], and a qualitative design was reported 3 times [63-65]. In 29 studies, a combination of services was offered; specifically, triage, testing, and a result service in 14 studies [13,28,40,42,46,49,51-53,56,57,59,60,63], triage and testing in 9 studies [26,27,29-33,35,37], and testing and a result service in 6 studies [41,44,45,48,61,64]. Furthermore, 8 studies discussed a testing service [14,25,34,38,43,47,58,62], 7 discussed a result service [15,35,39,50,54,55,65], and 1 discussed a triage service [12]. In the included studies, the testing service was evaluated most often (ie, 82% of the studies). Triage was evaluated in 2 studies [12,29] and the result service, in 11 studies [15,35,39-41,44,46,50,54,55,65]. The services were evaluated in the United States (n=15), the United Kingdom (n=9), Canada (n=6), Australia (n=2), Sweden (n=2), the Netherlands (n=2), and China (n=2). The remaining studies took place in Belgium, Brazil, Denmark, Estonia, France, Italy, and Uganda (ie, all n=1). The sample sizes ranged from 10 to 37 in the qualitative studies, with a mean of 21.60 (SD 9.7). The sample size ranged from 102 to 1736, with a mean of 2205.90 (SD 3514.0) in the quantitative studies. Almost half of the studies included both men and women (n=22) [12,13,25,29,36,38,39,48,50-57,59-62,64,65], 11 studies included MSM [27,28,34,35,41-43,45,47,49,63], 7 studies included only women [30-33,37,44,46], 2 studies included only men [26,58], 1 study included both MSM and transgender people [14], 1 study included adults with presumptive tuberculosis [15], and 1 study included past service users [40]. The mean percentage of male participants was 62.34% (SD 35.1%), and the mean age was 27.37 years (SD 4.7 years) (the average across the 15 studies that reported a mean) and ranged from 20.70 to 37.90 years. The study characteristics are shown in Table 1.

Table 1. Study characteristics.

Study and country	Study design	Study population	Sample size, n	Males, n (%)	Age (years)	Service type
Ahmed-Little et al, 2015 [61], the United Kingdom	Mixed methods	Persons aged ≥16 years	2247	1043 (46.41)	Mean 22.60	Testing[a], result
Andersen et al, 2001 [25], Denmark	Quantitative descriptive	Persons aged 21 to 23 years	183	64 (34.9)	—[b]	Testing
Babirye et al, 2019 [15], Uganda	Quantitative descriptive	Adults with presumptive tuberculosis	233	114 (48.9)	IQR 27-50	Result
Barnard, 2018 [51], the United Kingdom	Quantitative nonrandomized	Personsaged ≥16 years	5747	2489 (43.31)	IQR 23-32	Triage, testing[a], result
Brown, 2018 [56], the United Kingdom	Quantitative RCT[c]	High-risk persons aged ≥16 years	8999	7015 (77.95)	72% aged between 16 and 34	Triage, testing[a], result
Chai, 2010 [26], the United States	Quantitative descriptive	Men aged ≥14 years	501	501 (100.00)	IQR 21-30	Triage, testing[a]
de Boni, 2019 [27], Brazil	Quantitative descriptive	MSM[d] aged ≥18 years	3218	3218 (100.00)	IQR 22-31	Triage, testing[a]
Dulai, 2019 [49], Canada	Quantitative descriptive	Men who are gay, bisexual, and MSM aged ≥18 years	1272	1272 (100.00)	53% aged between 18 and 39	Triage, testing[a], result
Elliot, 2016 [28], the United Kingdom	Quantitative descriptive	MSM	17,361	17,361 (100.00)	—	Triage, testing[a], result
Grandahl et al, 2020 [64], Sweden	Qualitative	Personsaged ≥15 years	20	9 (45)	Mean 30.8	Testing[a], result
Grandahl, 2020 [48], Sweden	Quantitative descriptive	Personsaged ≥15 years	1785	546 (30.58)	Mean 27.3	Testing[a], result
Gaydos, 2016 [30], the United States	Quantitative descriptive	Women	102	0 (0)	64% aged between 18 and 29	Triage, testing[a]
Gaydos, 2016 [29], the United States	Quantitative descriptive	Persons aged ≥14 years	1394	558 (40.02)	Mean 28.13	Triage[a], testing
Gaydos, 2011 [32], the United States	Quantitative nonrandomized	Women aged ≥14 years	1171	0 (0.00)	Mean 25.00	Triage, testing[a]
Gaydos, 2009 [31], the United States	Quantitative descriptive	Women aged ≥14 years	1203	0 (0.00)	Median 23	Triage, testing[a]
Gaydos, 2006 [33], the United States	Quantitative descriptive	Women aged ≥14 years	400	0 (0.00)	Mean 26.10	Triage, testing[a]
Gilbert, 2019 [52], Canada	Quantitative nonrandomized	Persons aged ≥14 years	381	270 (70.86)	Range 18-74	Triage, testing[a], result[a]
Gilbert, 2017 [13], Canada	Quantitative descriptive	Persons aged ≥14 years	868	619 (71.31)	Median 32	Triage, testing[a], result
Jin, 2019 [34], China	Quantitative descriptive	MSM aged ≥16 years	879	879 (100.00)	IQR 24-34	Testing
Kersaudy-Rahib, 2017 [57], France	Quantitative RCT	Persons aged 18-24 years	11,075	5152 (46.52)	Mean 20.70	Triage, testing[a], result
Knight, 2018 [63], Canada	Qualitative	MSM aged ≥15 years	37	37 (100.00)	Mean 37.90	Triage, testing[a], result
Koekenbier, 2008 [35], the Netherlands	Quantitative descriptive	MSM	898	898 (100.00)	—	Result
Kuder, 2015 [53], the United States	Quantitative nonrandomized	Persons aged ≥14 years	1211	484 (39.97)	Mean 27.47	Triage, testing[a], result
Kwan, 2012 [36], Australia	Quantitative descriptive	Persons aged ≥16 years	377	206 (54.64)	71% were aged <30	Triage, testing[a]

Study and country	Study design	Study population	Sample size, n	Males, n (%)	Age (years)	Service type
Ladd, 2014 [37], the United States	Quantitative descriptive	Women	205	0 (0.00)	Mean 25.80	Triage, testing[a]
Ling, 2010 [54], the United States	Quantitative nonrandomized	Men and women	9056	5196 (57.37)	85% were aged ≥20	Result
Mák, 2015 [55], Canada	Quantitative nonrandomized	Persons aged ≥18 years	3292	1244 (37.79)	62% were aged ≥55	Result
Martin, 2009 [38], Australia	Quantitative descriptive	Persons aged 16-24 years	413	224 (54.2)	67% aged between 16 and 24	Testing
Morris, 2010 [39], the United States	Quantitative descriptive	Persons aged ≥18 years	3138	2563 (81.67)	62% aged between 25 and 44	Result
Nadarzynski, 2018 [40], the United Kingdom	Quantitative descriptive	Service users	115	—	—	Triage, testing, result[a]
Platteau, 2015 [41], Belgium	Quantitative descriptive	MSM aged ≥18 years	1071	1071 (100.00)	Mean 33.82	Testing, result[a]
Polilli, 2016 [12], Italy	Quantitative descriptive	Men and women	5000	—	—	Triage
Reagan, 2012 [58], the United States	Quantitative RCT	Men aged 18-45 years	200	200 (100.00)	Mean 30.75	Testing
Ricca, 2016 [42], the United States	Quantitative descriptive	MSM aged ≥18 years	896	896 (100.00)	Mean 30.00	Triage, testing[a], result
Robinson, 2019 [65], Canada	Qualitative	No inclusion criteria	21	12 (57)	38% aged between 60 and 69	Result
Rosengren, 2016 [43], the United States	Quantitative descriptive	Black and Hispanic MSM aged ≥18 years	125	125 (100.00)	63% aged between 18 and 30	Testing
Rotblatt, 2013 [44], the United States	Quantitative descriptive	Women aged 12 to 25 years	2659	0 (0.00)	Median 22.3	Testing[a], result[a]
Rüütel, 2015 [45], Estonia	Quantitative descriptive	MSM aged ≥18 years	265	265 (100.00)	53% were aged ≥30	Testing[a], result
Spielberg, 2014 [46], the United States	Quantitative descriptive	Women aged 18-30 years	217	217 (100)	Median 25	Triage, testing[a], result[a]
Talboom-Kamp, 2020 [50], the Netherlands	Quantitative descriptive	No inclusion criteria	354	—	—	Result
Wilson, 2019 [60], the United Kingdom	Quantitative RCT	Persons aged 16-30 years whom had never had a sexually transmitted infection test	528	254 (48.1)	Mean 21.30	Triage, testing[a], result
Wilson, 2017 [59], the United Kingdom	Quantitative RCT	Persons aged 16-30 years	2063	846 (41.01)	Mean 23.00	Triage, testing[a], result
Witzel, 2019 [14], the United Kingdom	Mixed methods	MSM and transgender people aged ≥16 years	1035/10	1035 (100)/ 10 (100)	IQR 26 to 42 or 60% aged between 26 and 40	Testing
Witzel, 2021[62], the United Kingdom	Mixed methods	Transgender people aged ≥16 years	118/20	94 (79.66)/ 12 (60)	IQR 22 to 37 or 35% aged between 16 and 25	Testing

Study and country	Study design	Study population	Sample size, n	Males, n (%)	Age (years)	Service type
Zhong, 2017 [47], China	Quantitative descriptive	MSM aged ≥18 years	380	380 (100)	54% aged between 25 and 34	Testing

[a]When multiple services were discussed in a study, footnote a identifies the service for which data was reported.

[b]—: data not available.

[c]RCT: randomized controlled trial.

[d]MSM: men who have sex with men.

Service Provider Characteristics

Within the 45 studies included in this review, 31 different providers were examined. The characteristics of the service providers are shown in Table 2, and more details are provided in Appendix 4 [12-15,25-65]. About half of the service providers offered a combination of services. A total of 9 providers offered a triage, testing, and result service, 5 offered a testing and result service, and 2 offered a triage and testing service. The remaining providers offered a single service (ie, testing [n=7], result [n=7], or triage [n=1]). Social marketing was most often used to recruit service users or study participants, with 16 providers using it as the sole recruitment strategy and 5 providers combining it with community outreach. The health service recruited 7 providers, and 3 studies reported no information on the applied recruitment strategy.

Triage was offered by 12 different service providers, either alone or in combination with other services. Triage aimed to estimate the risk of having a disease and identify individuals who need to test. The aim of the triage, however, was not specified for 5 providers. In most cases, web-based triage directed users to home-based testing (83%). A total of 23 providers offered testing as a service (alone or in combination with other services); 12 providers offered testing for 1 disease, and 11 offered testing for >2 diseases (ie, ranging from 2 to 6). Testing was most often available for chlamydia (n=13), HIV (n=12), and gonorrhea (n=10). Providers also tested for trichomonas (n=3), syphilis (n=3), hepatitis B (n=1), hepatitis C (n=1), lymphogranuloma venereum (n=1), and mycoplasmosis (n=1). Most of the tests were performed with a self-sampling test (n=18), whereby the samples were returned to the laboratory and analyzed according to the gold standard. All laboratories provided high-quality analysis with accredited and certified equipment. Self-testing was offered by 5 providers and targeted HIV (n=5) and syphilis (n=1). The testing service was almost always free of charge (87%). A small shipping fee was charged by 1 provider, and 1 provider charged US $23 that would be refunded after the user had shared the test results with the staff. A result service was offered by 20 providers (alone or in combination with other services). Different methods were used to communicate the test results, with 8 providers relying on a single method and 10 providers using different methods for result communication. Test results were most often accessible on the internet (n=12) or communicated over the phone (n=10). The results could also be communicated using SMS text messaging (n=6) or email (n=2). The communication of the test results was, in most cases, not completely independent from an HCP (70%). Often, the results were presented on the web, but users were called by the HCP when they had a positive result [39,63], or users were called when they had not checked their results on the internet [41].

Table 2. A description of the diagnostic testing and result service provider.

Service provider	Recruitment method[a]	Triage, type of follow-up testing	Testing			Result	
			Diseases	Type of home-based test	Cost on average (US $)	Method	Independent health care provider
Triage service							
Fai il test anche TU project [12]	Social	Clinic	HIV, hepatitis B and C, syphilis	__[b]	—	—	—
Testing service							
C-project [38]	Social	—	Chlamydia	Self-sampling	Free	—	—
Easy test [34]	Social; Community	—	HIV	Self-testing	2-3	—	—
UCLA free HIV self-test program [43]	Social	—	HIV	Self-testing	Free	—	—
Social entrepreneurship testing [47]	—	—	HIV, syphilis	Self-testing	23 (refunded)	—	—
SELPHI [14,62]	Social	—	HIV	Self-testing	Free	—	—
Unknown [25]	Social	—	Chlamydia	Self-sampling	Free	—	—
Unknown [58]	Social; Community	—	Chlamydia, gonorrhea	Self-sampling	Free	—	—
Unknown [48,64]	Health service	—	Chlamydia, gonorrhea	Self-sampling	Free	—	—
Result service							
GxAlert [15]	Health service	—	Tuberculosis	—	—	SMS text messaging	Yes
Syfilistest.nl [35]	Social	—	Syphilis	—	—	Web-based	Yes
Early test [39]	Social	—	HIV	—	—	Web-based; phone	Partly
Result system of Denver Metro Health Clinic [54]	Health service	—	Chlamydia, gonorrhea	—	—	Web-based	Partly
Excelleris [55]	Health service	—	Not limited to a specific disease	—	—	Web-based	Yes
Patient portal [50]	Health service	—	Not limited to a specific disease	—	—	Web-based	Partly
myCARE [65]	Health service	—	Not limited to a specific disease	—	—	Web-based	Partly
Triage and testing service							
A hora é Agora [27]	Social	Home	HIV	Self-testing	Free	—	—
Online Chlamydia Testing program [36]	Social	Home	Chlamydia, gonorrhea	Self-sampling	Free	—	—
Testing and result service							
Swab2Know [41]	Social	—	HIV	Self-sampling	Free	Web-based; email; phone	Partly
Do not think, know [44]	Social; Community	—	Chlamydia, gonorrhea	Self-sampling	Free	Web-based; phone	Partly

Service provider	Recruitment method[a]	Triage, type of follow-up testing	Testing			Result	
			Diseases	Type of home-based test	Cost on average (US $)	Method	Independent health care provider
Testikodus [45]	Social	—	Chlamydia, gonorrhea, trichomonas, LGV[c], mycoplasmosis	Self-sampling	Free	Web-based	Yes
RUClear [61]	—	—	HIV	Self-sampling	—	Phone; SMS text messaging; letter	Partly
Triage, testing, and result service							
DS@H [28]	Social	Home	HIV	Self-sampling	Free	SMS text messaging; web-based; phone	Partly
GetCheckedOnline [13,49,52,63][d]	Social	Home, clinic	Chlamydia, gonorrhea	Self-sampling	Free	Web-based; phone	Partly
Let's talk about it NHS [40]	Health service	Home	Chlamydia, gonorrhea, HIV, syphilis, hepatitis B and C	Self-sampling	Free	SMS text messaging; phone	Partly
Checking in [42]	Social	Home	HIV	Self-sampling	Free	Phone	Partly
eSTI [46]	Social; Community	Home	Chlamydia, gonorrhea, trichomonas	Self-sampling	Free	Web-based	Yes
SH:24 [48,59,60][d]	Social; Community	Home	Chlamydia, gonorrhea, HIV, syphilis	Self-sampling	Free	SMS text messaging; phone	Partly
Freetesting.hiv [56]	—	Home	HIV	Self-sampling	Free	SMS text messaging; phone	Partly
Chlamyweb [57]	Social	Home	Chlamydia	Self-sampling	Free	Email; postal service	Partly
I Want The Kit [26,29-33,37,53][d]	Social	Home	Chlamydia, gonorrhea, trichomonas	Self-sampling	Free	Web-based	Yes

[a]The methods used to recruit participants or service users was reported; specifically, social=social marketing, community=community outreach, and health service=health service recruitment.

[b]Data not available.

[c]Lymfogranuloma venereum.

[d]The service provider was investigated in multiple studies. The specific characteristics of each study are presented in Multimedia Appendix 3.

Quality Assessment

Quality assessment using the MMAT of the studies is shown in Table 3. The quality of the included studies was good, with an average score of 3.86 (SD 0.6; on a scale from 0 to 6). The average quality score ranged from 3.33 (SD 1.5) for mixed methods studies to 4.67 (SD 0.57) for qualitative studies. A shortcoming was that, in the studies using a quantitative descriptive design, the nonresponse was not clearly reported in 23 of the 25 studies. Therefore, it is unclear if these studies were at risk of nonresponse bias.

Table 3. Quality assessment of the included studies using the Mixed Method Appraisal Tool (MMAT).

Included studies	MMAT quality criteria[a]					MMAT scores[b]
	1	2	3	4	5	
Qualitative						4.67
Knight et al [63]	+[c]	+	+	+	+	5
Grandahl et al [64]	+	+	+	+	+	5
Robinson et al [65]	(+/−)[d]	+	+	+	+	4
Quantitative randomized controlled trials						4.20
Brown et al [56]	+	+	+	+/−	+	4
Kersaudy-Rahib et al [57]	+	+	_[e]	+/−	+	3
Reagan et al [58]	+	+	−	+	+	4
Wilson et al [59]	+	+	+	+	+	5
Wilson et al [60]	+	+	+	+	+	5
Quantitative nonrandomized						3.83
Gaydos et al [32]	+	+	+	+/−	+	4
Barnard et al [51]	+	+	−	+	+	4
Gilbert et al [52]	−	+	+/−	+	+	3
Kuder et al [53]	+	+	−	−	+	3
Ling et al [54]	+	+	+	+	+	5
Mák et al [55]	−	+	+	+	+	4
Quantitative descriptive						3.78
Polilli et al [12]	+	+	+	+/−	+	4
Gilbert et al [13]	+	+	+	+/−	+	4
Babirye et al [15]	+	+	+	+	+	5
Andersen et al [25]	+	+	+	+/−	+	4
Chai et al [26]	+	+	+	+/−	+	4
de Boni et al [27]	+	+	+	+/−	+	4
Elliot et al [28]	+	+	+	+/−	+/−	3
Gaydos et al [29]	+	+	+	+/−	+	4
Gaydos et al [30]	+	+	+	+/−	+	4
Gaydos et al [31]	+	+	+	+/−	+	4
Gaydos et al [33]	+	+	+	+/−	+	4
Jin et al [34]	+	+	+	+/−	+	4
Koekenbier et al [35]	+	+	+	+/−	+	4
Kwan et al [36]	+	−	+	+/−	+	3
Ladd et al [37]	+	+	+	+/−	+	4
Martin et al [38]	+	−	+	+/−	+	3
Morris et al [39]	+	+	+	+/−	+	4
Nadarzynski et al [40]	+	+/−	+	+/−	+	3
Platteau et al [41]	+	+	−	+/−	+	3
Ricca et al [42]	+	+	+	+/−	+	4
Rosengren et al [43]	+	+	+	+/−	+	4
Rotblatt et al [44]	+	+	+	+/−	+	4

Included studies	MMAT quality criteria[a]					MMAT scores[b]
	1	2	3	4	5	
Rüütel et al [45]	+	−	+	−	+	3
Spielberg et al [46]	+	+	+	+/−	+	4
Zhong et al [47]	+/−	+	+	+/−	+	3
Grandahl et al [48]	+	+	+	−	+	4
Dulai et al [49]	+	+	+	−	+	4
Talboom-Kamp et al [50]	+	+	+	−	+	4
Mixed methods						3.33
Witzel et al [14]	+	+	+	+	−	4
Ahmed-Little et al [61]	+/−	−	+	+	−	2
Witzel et al [62]	+	+	+	+/−	+	4

[a]The criteria differed according to the design. A description of the criteria is provided in Multimedia Appendix 2.

[b]The average Mixed Method Appraisal Tool score across all designs is 3.86. The overall grade is the sum of the number of quality criteria that were assessed as good.

[c]Good quality.

[d]Insufficient evidence to determine the quality.

[e]Poor quality.

Findings by Type of Service

Overview

The findings are discussed separately for triage, testing, and result service. For clarity, the findings of follow-up testing and treatment are jointly discussed for the testing and result service. A more detailed description of the findings is provided in Multimedia Appendix 4.

Triage Service

A total of 2 studies evaluated the triage service, which showed that the use of web-based triage services could be quite high with those completing the web-based triage and booking an appointment for a test (more than 50%). Notably, most of the individuals who tested positive were also linked to treatment. Furthermore, the predictive value of triage showed a prediction of STI positivity in women. For more detailed information, see Table 4.

Table 4. Results of the triage and test services per specific outcome measure.

Service and general outcome	Specific outcome measure	Results
Triage		
__[a]		
	Use	• Use of web-based triage services can be quite high; more than 50% (3046/5000) of those who completed the web-based triage also booked an appointment for HIV clinic-based testing. Notably, the majority also presented for testing (87%), and most of the individuals who tested positive were also linked to treatment (93%) [12]
	Predictive value	• Gaydos et al [29] found that the score on the risk assessment predicted STI[b] positivity for females but not males
Test		
Use		
	Return specimen	• The percentage of returned tests or specimens for analyses was frequently reported [13,25,26,28,37,38,42,44-46,48,51,56,61] • Range: 24 [45] to 85% [42,48]; mean 52.8% (SD 19.6%)
	Used tests	• In 4 studies, the percentage of used home-based tests was given [14,36,43,47] • Range: 56 [36] to 100% [43]; mean 83% (SD 19.3%) • The highest percentage might be an overestimation of the actual use because people had to self-report the use of the tests in a follow-up survey [43]
	Comparison home-based testing vs clinic-based testing	• In 4 studies, home-based testing was compared with clinic testing [57-60] • The average percentage of test use was higher among those who were offered a home test compared with those who were offered a test at the clinic (mean 49%, SD 17.8% vs mean 27%, SD 16.1%, respectively)
	Other	• Home-based test uptake was highest when the results would be presented through the internet [53] • When users received primers before the arrival of the test kit at home (eg, set aside a time to complete the test) and behavioral insight reminders [56]
Acceptability or usability		
	Home-based testing vs clinic-based testing	• Eight studies examined whether there was a preference for home-based or clinic-based testing [26,30,32,33,43,46,63] • Range: 62 [30] to 95% [46]; mean 81% (SD 12.7%) who preferred home-based testing • One study reported a barrier to clinic-based testing: that it was easier to stay at home than go to the clinic [49]
	Easy to perform	• Seven studies reported how easy it was to perform home-based testing [14,26,30,32,33,36,43] • Range: 88% [26] to 97% [14,32]; mean 94% (SD 3.5%)
	Acceptability instructions	• Five studies examined the acceptability of the instructions for home-based testing [14,27,30,58,61] • Mean 93% (SD 5.3%) considered the instructions to be easy.
	Acceptability in general	• In 3 studies, the acceptability of the home-based test service, in general, was reported [59-61] • Mean 75% (SD 4.5%)
	Recommendation	The percentage of participants who would recommend the service of testing at home to a friend was 98% in 2 studies [36,46], and in Gaydos et al [30], it was 77%

Service and general outcome	Specific outcome measure	Results
	Other	• The perceived reliability of the test results was reported in Gaydos et al [30]: 97% of the users trusted the results of the home-based test service • Chai et al [26] found that 85% found it a safe way of testing • Witzel et al [14] found that 97% had an overall good experience with the home-based test service • Chai et al [26], Gaydos et al [32], and Dulai et al [49] both reported that around 90% would use the home-based test service again • Gaydos et al [33] report that 86% would use this home-based testing method in daily life • de Boni et al [27] reported that 91% found it (very) easy to use the website • Grandahl et al [48] reported that more than 90% found the overall home-based test service good or very good • Grandahl et al [64] reported that most users highly appreciated the service and found the service easy to use, convenient, and confidential. They would use the service again in the future, even if the costs were higher
Cost-effectiveness	Cost-effectiveness	• Kersaudy-Rahib et al [57] reported that the price for home-based testing was three times lower compared with clinic-based testing • Ahmed-Little et al [61] showed that the costs for HIV testing per person were around “€27 (US $ 30.45), which is in line with testing costs in national HIV testing pilots
Other outcomes	—	• The reasons to self-test were that it reduced HIV testing barriers, desire to use new technology, and altruistic motivation [14] • Other reasons mentioned for HIV self-testing were inaccessible and inappropriate clinical services [62]. In Martin et al [38] users reported that they did the test because it was easy and it was for free • Zhong et al [47] reported convenience and to save time, protection of privacy, ease of use, and accuracy as reasons to perform a home-based self-test. Facilitators were ease of use, anonymity, and the ability to test alone. Barriers were concerns about accuracy, potential costs, and concerns about self-interpreting the results • Dulai et al [49] reported that 20% were worried about their online information privacy, and 5% had low trust in this service • Some barriers mentioned in Grandahl et al [64] were the use of complicated language, uncertainty about the procedure, unreliable postal service, and insecure data handling

[a]No general outcome measure.

[b]STI: sexually transmitted infection.

Testing Service

For the test service, different outcome measures were found with different objectives. Studies with outcomes focusing on the test services, which were home-based (eg, self-testing or self-sampling), were discussed. The test use was reported to be high (above 50%), and test uptake was higher among those offered home-based tests than clinic-based tests. The number of returned specimens was discussed frequently and showed very different results with a wide range of percentages of returned specimens. The acceptability and usability of the test service scored high on the convenience of performing home-based tests with easy instructions. The cost-effectiveness of home-based tests showed lower or similar prices compared with clinic-based testing. Furthermore, motivations for self-testing were discussed. Ease of use, privacy, and anonymity were identified as reasons to perform these tests. Important barriers for these services were potential costs, accuracy, unreliable postal service, insecurity about handling data, and self-interpreting the results. For more detailed information, see Table 4.

Result Service

For the result service, different types of outcome measures were found with different objectives. The use of the result service exceeded 69%. Research showed that most participants viewed their results on the same day as they were posted on the web, and comprehension of these web-based results was high (above 75%). The acceptability of direct access to results using the website was high, and the participants were satisfied with this process. Direct access to diagnostic results led to shorter waiting times for the results than for participants who did not receive their results on the web. Limited access to the internet was a reason for preferring to call the clinic for the results. For more detailed information, see Table 5.

Table 5. Results of the test and result services per specific outcome measure.

Service and general outcome	Specific outcome measure	Results
Result		
Use		
	Retrieved results on the internet	• The use of a result service was assessed in 6 studies [35,39,41,44,46,54] • The percentage of people who retrieved their results on the internet varied from 69 [39] to 97% [35]; mean 85% (SD 11.2%) • The service with the lowest retrieval rate called all users with a positive test result and, if users were not called within 2 week they could access their results on the internet • Spielberg et al [46] found that 88% viewed their test results on the same day that the results were posted • Platteau et al [41] showed that significantly more people collected their test results when the test was ordered online compared with testing during outreach activities
	Waiting time	• Gilbert et al [52] showed significantly shorter waiting times for those who used a web-based platform compared with clinic clients
Comprehension	__[a]	• Babirye et al [15] found that everyone could accurately relay the content of an SMS text message that contained the tuberculosis test result • Comprehension was slightly lower in the other 2 studies: 75% and 87% understood the content of the test result message, respectively [40,55] • Mák et al [55] showed that comprehension was significantly higher in the group that did not receive their results on the internet • Robinson et al [65] showed that comprehension of the results differed from difficulty with the understanding of the results to no difficulty. However, when difficulties were there, the users pointed out that the reference range was helpful.
Acceptability		
	Comfortable with web-based results	• The acceptability was examined in 4 different studies [39,41,46,54] • Only 1 study specifically examined how comfortable users were with receiving their results on the internet, and 87% was (very) comfortable with this process [39]
	Ordering a test and receiving results on the web	• Two studies examined the acceptability of ordering a test kit on the web and receiving the web-based results • Platteau et al [41] found that 96% of the users were satisfied with this process • Spielberg et al [46] reported that 98% of the users found the service website easy to use
	Reasons	• The two main reasons for choosing to receive web-based results were having access to the results any time of the day and the belief that results would be communicated faster via the internet • A preference to call the clinic for results and limited access to the internet were reasons to opt-out of web-based results [54] • The reasons for using web-based results were reported by Robinson et al [65] as better communication with the HCP[b], convenience, and being a steward of your health care

Service and general outcome	Specific outcome measure	Results
Other outcomes	—	• The feasibility of using SMS text message to communicate tuberculosis test results was examined in Uganda and scored relatively low; (ie, an SMS text message was only transmitted to 62% of those who were eligible to receive an SMS text message with test results [15]) • One study found that users waited significantly shorter for web-based test results than users who did not have web-based access [55]. Furthermore, this study showed that the majority (ie, 86%) experienced no or low anxiety after receiving their test results, and the level of anxiety was not different between those with or without internet access • Another study examined user preferences for the content of the text messages conveying the test results, and the majority preferred that the results of all tested STIs[c] were discussed in one message and that the names of the STIs tested should be included in the message [40] • One study reported that patients feel more comfortable and engaged with their health care when they see the results themselves [65]. Besides, they reported that it had no adverse effects • Two domains of the eHIQ[d] were researched in one study to determine patient's attitude toward a web-based results service [50]. This eHIQ showed positive results for the criteria: easy to use, trustworthy, and appropriate
Test and result		
Follow-up testing and treatment		
	Confirmatory testing	• The frequency of confirmatory testing for positive or uncertain or invalid test results was described in 4 studies [27,35,43,61] • Range from 68% [27] to 100% [43,61]; mean 85% (SD 17.7%)
	Follow-up after positive result	• Follow-up treatment after a positive test result was described in 10 studies [26,31,32,34,36,41-44,46] • Receiving web-based test results led to high treatment rates; mean 93% (SD 9.9%)
	Confirmatory testing and treatment	• In 2 studies, confirmatory testing and treatment were described [28,47] • In Elliot et al [28], 67% of the reactive samples were confirmed, and all received treatment. For 10% of the reactive samples, treatment could not be confirmed • In Zhong et al [47], everyone with a reactive test did confirmatory testing and was linked to treatment
	Other	In 3 studies, different groups were compared with each other. It was shown that the treatment rate was higher when users (1) had the option to receive web-based results versus communicated over the phone (not significant) [54], (2) received their test kit at home instead of at the primary care setting [57], and (3) received their results through an automated result access system compared with service where participants had to call for their test result [53]

[a]Data not available.

[b]HCP: health care professional.

[c]STI: sexually transmitted infection.

[d]eHIQ: e-Health Impact Questionnaire.

Test and Result Services: Follow-Up Testing and Treatment

Follow-up testing and treatment have been discussed in several studies. These studies showed that receiving web-based results led to high treatment rates (mean 93%, SD 9.9%), and the frequency of confirmatory testing after a self-test was above 68%. For more details, see Table 5.

Discussion

Principal Findings

This systematic review aimed to gain insight into the available methods for direct web-based access to patients for diagnostic testing and results. A total of 45 studies were included. Most of the studies used a quantitative descriptive design. Most of the studies investigated a test or result service related to STIs. In the 45 studies, 31 different providers were discussed. Half of the providers offered a combination of services. Of the 3 different services, the test service was most often evaluated. This review showed that direct patient access to testing and result services was positively evaluated. The use of triage, test,

and result services was high, and the acceptability among patients was high. Moreover, follow-up confirmatory testing and treatment rates were high with home-based testing.

An update of the literature search was performed after the third wave of the COVID-19 pandemic. However, no studies were found regarding direct access to diagnostic testing and results services for this disease. This could be because free tests were often offered by the governments of countries. There have been commercial companies offering tests for SARS-CoV-2; however, scientific research has not yet been performed.

This review found that the use rates of home-based tests were high and that direct web-based access to results was appreciated and generally well-understood. An overall preference for home-based testing versus clinic-based testing was found. Importantly, follow-up treatment after a positive home-based test was high and, in some studies, was even higher when tests were performed at home compared with the clinic. The overall positive findings of this systematic review contradict earlier voiced concerns about self-testing and self-sampling, such as that users would be insufficiently linked to follow-up testing or treatment [66,67]. It was reported in 1 study that 70% of participants were afraid to carry out the self-test properly [67]. This contrasted with our findings, which indicated that users found self-tests easy to use and that the instructions were clear and reliable. Nevertheless, it is important to include end users in the design phase when setting up such services to ensure usability and acceptability [68]. In addition, although most studies reported high acceptability and comprehension of test results communicated on the web, 1 study reported that interpreting the results was easier when they were communicated in person (vs via the internet). This contradictory finding might be because this study discussed a general result service portal and not a portal specifically for STI results. To minimize the risk of misunderstanding, it is important that future research examine the content and how this content can best be presented to users [50].

Furthermore, the quality of the laboratory tests used in these studies was high. Therefore, this review disproves the aforementioned concerns about home-based diagnostic tests [66,67] and shows that these tests with direct access to web-based result services could contribute to easily accessible diagnostic testing [69].

The high acceptability of the test and result services and the high rates of follow-up for treatment create opportunities for primary care. The workload for primary care is high [3,4]. eHealth technologies can make health care delivery more efficient, and therefore, the adoption of eHealth is being stimulated worldwide [9]. By providing patients with direct access to web-based testing and results, patients would not need to visit their HCP, potentially lowering the number of consultations in primary care. Consequently, it would leave HCPs with more time to focus on complex health care and consultations that cannot be executed via the internet. Another reason for home-based diagnostic testing is to lower the testing threshold. Patients can experience feelings of embarrassment or shame for tests such as STI, which can result in delays in testing [70]. Allowing individuals to order tests on the web can make it more convenient for them to get tested and may help diagnose and treat diseases sooner. However, future research should investigate whether these types of test services lead to excessive use. At the same time, it is important to emphasize that this review identified that direct access to diagnostic testing exhibited benefits for patients, such as comfort, ease, and time-saving. A few barriers should be addressed to allow home-based diagnostic testing in practice. An important barrier to eHealth adoption in primary care is, for example, the cost [71]. In the Netherlands, diagnostic tests ordered by a primary care physician are covered by health insurance. However, home-based diagnostic testing has not yet been covered by insurance. To stimulate home-based testing, the costs of home-based diagnostic testing should be covered by an individual's health care insurance. Therefore, it would be useful to investigate the cost-effectiveness of home-based diagnostic testing compared with clinic-based testing. In this review, only 2 studies discussed cost-effectiveness, more insight into how valuable home-based diagnostic testing could be in the future could be provided. Furthermore, home-based diagnostic testing could work more efficiently in primary care if implemented for a variety of conditions [72]. However, more research is needed to elaborate on home-based diagnostic test services for diseases other than STIs.

Strengths and Limitations

The strengths of this review lie in several aspects. First, the study search strategy was comprehensive and not limited to a specific disease or population. Second, a quality assessment was performed for all included studies, and the quality of the included studies appeared to be relatively high. However, it is essential to consider that the MMAT was scored using a yes or no score without nuances. Third, a comprehensive overview of the study and service characteristics provided detailed insight into the included studies.

This review has several limitations. First, there was heterogeneity in the included outcome measures, which resulted in a low number of studies reporting the same outcome. Therefore, it was not possible to examine the pooled effect using a meta-analysis. As the field advances quickly, more studies are likely to become available soon, and a meta-analysis might be possible. Second, almost all studies focused on STIs. For that reason, it was unknown whether the findings regarding usability and acceptability would generalize to test and result services that target diseases other than STIs. Nevertheless, our review provided insight into the potential of direct web-based access to diagnostic testing, which could translate to other diseases. Even for test results that were not dichotomous, which was the case in STI testing, test results could be presented in a web-based portal, for example, the identification of abnormal and normal values for a test result with an option to contact a physician [50]. A third limitation was that the mean age in the included studies was relatively low, which could have led to bias because a different, older population could have evaluated these services differently [73]. Although eHealth services have shown good use and result in older adult populations, it remains to be determined whether this is also the case for web-based diagnostic testing and results services [74]. There was a large portion of the quantitative descriptive design studies (28/45,

62%) that constituted the fourth limitation to this review. Only 5 studies had a randomized controlled trial design. Therefore, selection bias cannot be ruled out, including sample representativeness. Nevertheless, all studies underwent quality assessment and scored relatively high.

Conclusions

Home-based testing showed higher use rates and follow-up treatment rates compared with clinic-based testing. It was demonstrated to be acceptable, safe, and convenient for users, which could lower the threshold for testing. Future research on diagnostic testing for diseases other than STIs and cost-effectiveness evaluation is needed. To conclude, this review showed that eHealth technologies for diagnostic testing could contribute to easy direct access to high-quality diagnostic testing for patients and has the potential to increase efficiency and possibility to reduce workload in primary care. In conclusion, direct web-based access to diagnostic testing showed promising results.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search terms for this systematic review.
[DOCX File , 21 KB - jmir_v24i1e29303_app1.docx]

Multimedia Appendix 2
Mixed Method Appraisal Tool.
[DOCX File , 14 KB - jmir_v24i1e29303_app2.docx]

Multimedia Appendix 3
Server provider characteristics per study.
[DOCX File , 29 KB - jmir_v24i1e29303_app3.docx]

Multimedia Appendix 4
Overview of the reported outcomes for the triage, testing, and result services.
[DOCX File , 36 KB - jmir_v24i1e29303_app4.docx]

References

1. Prince MJ, Wu F, Guo Y, Gutierrez Robledo LM, O'Donnell M, Sullivan R, et al. The burden of disease in older people and implications for health policy and practice. Lancet 2015 Feb;385(9967):549-562. [doi: 10.1016/s0140-6736(14)61347-7]
2. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. Lancet 2012 Jul;380(9836):37-43. [doi: 10.1016/s0140-6736(12)60240-2]
3. Konrad T, Link C, Shackelton R, Marceau L, von dem Knesebeck O, Siegrist J, et al. It's about time: physicians' perceptions of time constraints in primary care medical practice in three national healthcare systems. Med Care 2010 Mar;48(2):95-100 [FREE Full text] [doi: 10.1097/MLR.0b013e3181c12e6a] [Medline: 20057331]
4. Mohr D, Benzer J, Young G. Provider workload and quality of care in primary care settings: moderating role of relational climate. Med Care 2013 Jan;51(1):108-114. [doi: 10.1097/MLR.0b013e318277f1cb] [Medline: 23222471]
5. van den Hombergh P, Künzi B, Elwyn G, van Doremalen J, Akkermans R, Grol R, et al. High workload and job stress are associated with lower practice performance in general practice: an observational study in 239 general practices in the Netherlands. BMC Health Serv Res 2009 Jul 15;9(1):118 [FREE Full text] [doi: 10.1186/1472-6963-9-118] [Medline: 19604386]
6. Groenewegen PP, Hutten JB. Workload and job satisfaction among general practitioners: a review of the literature. Soc Sci Med 1991 Jan;32(10):1111-1119. [doi: 10.1016/0277-9536(91)90087-s]
7. Goetz K, Musselmann B, Szecsenyi J, Joos S. The influence of workload and health behavior on job satisfaction of general practitioners. Fam Med 2013 Feb;45(2):95-101 [FREE Full text] [Medline: 23378076]
8. Stoffers J. The promise of eHealth for primary care: opportunities for service delivery, patient-doctor communication, self-management, shared decision making and research. Eur J Gen Pract 2018 Dec 26;24(1):146-148 [FREE Full text] [doi: 10.1080/13814788.2018.1449779] [Medline: 29577773]
9. van der Kleij RM, Kasteleyn MJ, Meijer E, Bonten TN, Houwink EJ, Teichert M, et al. SERIES: eHealth in primary care. Part 1: concepts, conditions and challenges. Eur J Gen Pract 2019 Oct 10;25(4):179-189 [FREE Full text] [doi: 10.1080/13814788.2019.1658190] [Medline: 31597502]
10. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun 18;3(2):e20. [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]

11. Shaw T, McGregor D, Brunner M, Keep M, Janssen A, Barnet S. What is eHealth (6)? Development of a conceptual model for eHealth: qualitative study with key informants. J Med Internet Res 2017 Oct 24;19(10):e324 [FREE Full text] [doi: 10.2196/jmir.8106] [Medline: 29066429]
12. Polilli E, Sozio F, Di Stefano P, Sciacca A, Ursini T, Paoloni M, et al. Web-based HIV testing in Abruzzo, Italy: analysis of 15-month activity results. AIDS Patient Care STDS 2016 Oct;30(10):471-475. [doi: 10.1089/apc.2016.0082] [Medline: 27749107]
13. Gilbert M, Salway T, Haag D, Fairley CK, Wong J, Grennan T, et al. Use of GetCheckedOnline, a comprehensive web-based testing service for sexually transmitted and blood-borne infections. J Med Internet Res 2017 Mar 20;19(3):e81 [FREE Full text] [doi: 10.2196/jmir.7097] [Medline: 28320690]
14. Witzel TC, Gabriel MM, McCabe L, Weatherburn P, Gafos M, Speakman A, et al. Pilot phase of an internet-based RCT of HIVST targeting MSM and transgender people in England and Wales: advertising strategies and acceptability of the intervention. BMC Infect Dis 2019 Aug 08;19(1):699 [FREE Full text] [doi: 10.1186/s12879-019-4247-1] [Medline: 31391003]
15. Babirye D, Shete PB, Farr K, Nalugwa T, Ojok C, Nantale M, et al. Feasibility of a short message service (SMS) intervention to deliver tuberculosis testing results in peri-urban and rural Uganda. J Clin Tuberc Other Mycobact Dis 2019 Aug;16:100110 [FREE Full text] [doi: 10.1016/j.jctube.2019.100110] [Medline: 31720434]
16. Johnson CC, Kennedy C, Fonner V, Siegfried N, Figueroa C, Dalal S, et al. Examining the effects of HIV self-testing compared to standard HIV testing services: a systematic review and meta-analysis. J Int AIDS Society 2017 May 15;20(1):21594. [doi: 10.7448/ias.20.1.21594]
17. Figueroa C, Johnson C, Verster A, Baggaley R. Attitudes and acceptability on HIV self-testing among key populations: a literature review. AIDS Behav 2015 Nov 9;19(11):1949-1965 [FREE Full text] [doi: 10.1007/s10461-015-1097-8] [Medline: 26054390]
18. LeGrand S, Muessig K, Horvath K, Rosengren A, Hightow-Weidman L. Using technology to support HIV self-testing among MSM. Curr Opin HIV AIDS 2017 Sep;12(5):425-431 [FREE Full text] [doi: 10.1097/COH.0000000000000400] [Medline: 28617712]
19. Noble M, Jones AM, Bowles K, DiNenno EA, Tregear SJ. HIV testing among internet-using MSM in the United States: systematic review. AIDS Behav 2017 Feb 6;21(2):561-575 [FREE Full text] [doi: 10.1007/s10461-016-1506-7] [Medline: 27498198]
20. Jamil MS, Hocking JS, Bauer HM, Ali H, Wand H, Smith K, et al. Home-based chlamydia and gonorrhoea screening: a systematic review of strategies and outcomes. BMC Public Health 2013 Mar 04;13(1):189 [FREE Full text] [doi: 10.1186/1471-2458-13-189] [Medline: 23496833]
21. Pai NP, Sharma J, Shivkumar S, Pillay S, Vadnais C, Joseph L, et al. Supervised and unsupervised self-testing for HIV in high- and low-risk populations: a systematic review. PLoS Med 2013 Apr 2;10(4):e1001414 [FREE Full text] [doi: 10.1371/journal.pmed.1001414] [Medline: 23565066]
22. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
23. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. The Mixed Methods Appraisal Tool (MMAT) version 2018 for information professionals and researchers. Edu Inform 2018 Dec 18;34(4):285-291. [doi: 10.3233/efi-180221]
24. McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012:276-282. [doi: 10.11613/bm.2012.031]
25. Andersen B, Østergaard L, Møller JK, Olesen F. Effectiveness of a mass media campaign to recruit young adults for testing of Chlamydia trachomatis by use of home obtained and mailed samples. Sex Transm Infect 2001 Dec;77(6):416-418 [FREE Full text] [doi: 10.1136/sti.77.6.416] [Medline: 11714938]
26. Chai S, Aumakhan B, Barnes M, Jett-Goheen M, Quinn N, Agreda P, et al. Internet-based screening for sexually transmitted infections to reach nonclinic populations in the community: risk factors for infection in men. Sex Transm Dis 2010 Dec;37(12):756-763 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181e3d771] [Medline: 20644498]
27. De Boni RB, Veloso VG, Fernandes NM, Lessa F, Corrêa RG, Lima RD, et al. An internet-based HIV self-testing program to increase HIV testing uptake among men who have sex with men in Brazil: descriptive cross-sectional analysis. J Med Internet Res 2019 Aug 01;21(8):e14145 [FREE Full text] [doi: 10.2196/14145] [Medline: 31373276]
28. Elliot E, Rossi M, McCormack S, McOwan A. Identifying undiagnosed HIV in men who have sex with men (MSM) by offering HIV home sampling via online gay social media: a service evaluation. Sex Transm Infect 2016 Sep 24;92(6):470-473. [doi: 10.1136/sextrans-2015-052090] [Medline: 27221475]
29. Gaydos CA, Jett-Goheen M, Barnes M, Dize L, Barnes P, Hsieh Y. Use of a risk quiz to predict infection for sexually transmitted infections: a retrospective analysis of acceptability and positivity. Sex Transm Infect 2016 Feb 18;92(1):44-48 [FREE Full text] [doi: 10.1136/sextrans-2015-052058] [Medline: 26285773]
30. Gaydos CA, Jett-Goheen M, Barnes M, Dize L, Hsieh Y. Self-testing for Trichomonas vaginalis at home using a point-of-care test by women who request kits via the internet. Sex Health 2016;13(5):491. [doi: 10.1071/sh16049]

31. Gaydos C, Barnes M, Aumakhan B, Quinn N, Agreda P, Whittle P, et al. Can e-technology through the internet be used as a new tool to address the Chlamydia trachomatis epidemic by home sampling and vaginal swabs? Sex Transm Dis 2009 Sep;36(9):577-580 [FREE Full text] [doi: 10.1097/OLQ.0b013e3181a7482f] [Medline: 19543145]
32. Gaydos C, Barnes M, Aumakhan B, Quinn N, Wright C, Agreda P, et al. Chlamydia trachomatis age-specific prevalence in women who used an internet-based self-screening program compared to women who were screened in family planning clinics. Sex Transm Dis 2011 Mar;38(2):74-78 [FREE Full text] [doi: 10.1097/OLQ.0b013e3182039d7f] [Medline: 21173720]
33. Gaydos C, Dwyer K, Barnes M, Rizzo-Price P, Wood B, Flemming T, et al. Internet-based screening for Chlamydia trachomatis to reach non-clinic populations with mailed self-administered vaginal swabs. Sex Transm Dis 2006 Jul;33(7):451-457. [doi: 10.1097/01.olq.0000200497.14326.fb] [Medline: 16652069]
34. Jin X, Xu J, Smith MK, Xiao D, Rapheal ER, Xiu X, et al. An internet-based self-testing model (Easy Test): cross-sectional survey targeting men who have sex with men who never tested for HIV in 14 provinces of China. J Med Internet Res 2019 May 15;21(5):e11854 [FREE Full text] [doi: 10.2196/11854] [Medline: 31094339]
35. Koekenbier R, Davidovich U, van Leent EJ, Thiesbrummel H, Fennema H. Online-mediated syphilis testing: feasibility, efficacy, and usage. Sex Transm Dis 2008 Aug;35(8):764-769. [doi: 10.1097/OLQ.0b013e31816fcb0a] [Medline: 18496472]
36. Kwan KS, Jachimowicz EA, Bastian L, Marshall L, Mak DB. Online chlamydia testing: an innovative approach that appeals to young people. Med J Aust 2012 Sep 03;197(5):287-290. [doi: 10.5694/mja11.11517] [Medline: 22938127]
37. Ladd J, Hsieh Y, Barnes M, Quinn N, Jett-Goheen M, Gaydos CA. Female users of internet-based screening for rectal STIs: descriptive statistics and correlates of positivity. Sex Transm Infect 2014 Sep 06;90(6):485-490 [FREE Full text] [doi: 10.1136/sextrans-2013-051167] [Medline: 24604333]
38. Martin L, Freedman E, Burton L, Rutter S, Knight V, D'Amato A, et al. The C-project: use of self-collection kits to screen for Chlamydia trachomatis in young people in a community-based health promotion project. Sex Health 2009;6(2):157. [doi: 10.1071/sh08039]
39. Morris SR, Little SJ, Cunningham T, Garfein RS, Richman DD, Smith DM. Evaluation of an HIV nucleic acid testing program with automated Internet and voicemail systems to deliver results. Ann Intern Med 2010 Jun 15;152(12):778-785 [FREE Full text] [doi: 10.7326/0003-4819-152-12-201006150-00005] [Medline: 20547906]
40. Nadarzynski T, Scholfield C, Symonds Y, Graham C, Kidsley S. Preferences for the format of text messages containing results of online screening for sexually transmitted infections: a service evaluation. Int J STD AIDS 2018 Sep 13;29(10):1014-1016. [doi: 10.1177/0956462418765274] [Medline: 29652223]
41. Platteau T, Fransen K, Apers L, Kenyon C, Albers L, Vermoesen T, et al. Swab2know: an HIV-testing strategy using oral fluid samples and online communication of test results for men who have sex with men in Belgium. J Med Internet Res 2015 Sep 01;17(9):e213 [FREE Full text] [doi: 10.2196/jmir.4384] [Medline: 26330138]
42. Ricca AV, Hall EW, Khosropour CM, Sullivan PS. Factors associated with returning at-home specimen collection kits for HIV testing among internet-using men who have sex with men. J Int Assoc Provid AIDS Care 2016 Nov 20;15(6):463-469 [FREE Full text] [doi: 10.1177/2325957416668579] [Medline: 27635015]
43. Rosengren AL, Huang E, Daniels J, Young SD, Marlin RW, Klausner JD. Feasibility of using GrindrTM to distribute HIV self-test kits to men who have sex with men in Los Angeles, California. Sex Health 2016;13(4):389. [doi: 10.1071/sh15236]
44. Rotblatt H, Montoya JA, Plant A, Guerry S, Kerndt PR. There's no place like home: first-year use of the "I know" home testing program for Chlamydia and Gonorrhea. Am J Public Health 2013 Aug;103(8):1376-1380. [doi: 10.2105/ajph.2012.301010]
45. Rüütel K, Lohmus L, Jänes J. Internet-based recruitment system for HIV and STI screening for men who have sex with men in Estonia, 2013: analysis of preliminary outcomes. Eurosurveillance 2015;20(15-16):21094. [doi: 10.2807/1560-7917.es2015.20.15.21094]
46. Spielberg F, Levy V, Lensing S, Chattopadhyay I, Venkatasubramanian L, Acevedo N, et al. Fully integrated e-services for prevention, diagnosis, and treatment of sexually transmitted infections: results of a 4-county study in California. Am J Public Health 2014 Dec;104(12):2313-2320. [doi: 10.2105/ajph.2014.302302]
47. Zhong F, Tang W, Cheng W, Lin P, Wu Q, Cai Y, et al. Acceptability and feasibility of a social entrepreneurship testing model to promote HIV self-testing and linkage to care among men who have sex with men. HIV Med 2017 May 07;18(5):376-382 [FREE Full text] [doi: 10.1111/hiv.12437] [Medline: 27601301]
48. Grandahl M, Mohammad J, Larsson M, Herrmann B. Users' opinions of internet-based self-sampling tests for Chlamydia trachomatis and Neisseria gonorrhoeae in Sweden. Acta Derm Venereol 2020 Nov 04;100(18):adv00315 [FREE Full text] [doi: 10.2340/00015555-3677] [Medline: 33104232]
49. Dulai J, Salway T, Thomson K, Haag D, Lachowsky N, Grace D, et al. Awareness of and intention to use an online sexually transmitted and blood-borne infection testing service among gay and bisexual men in British Columbia, two years after implementation. Can J Public Health 2021 Feb 16;112(1):78-88 [FREE Full text] [doi: 10.17269/s41997-020-00323-4] [Medline: 32557285]
50. Talboom-Kamp E, Tossaint-Schoenmakers R, Goedhart A, Versluis A, Kasteleyn M. Patients' attitudes toward an online patient portal for communicating laboratory test results: real-world study using the eHealth impact questionnaire. JMIR Form Res 2020 Mar 04;4(3):e17060 [FREE Full text] [doi: 10.2196/17060] [Medline: 32024632]

51. Barnard S, Free C, Bakolis I, Turner KM, Looker KJ, Baraitser P. Comparing the characteristics of users of an online service for STI self-sampling with clinic service users: a cross-sectional analysis. Sex Transm Infect 2018 Aug 07;94(5):377-383 [FREE Full text] [doi: 10.1136/sextrans-2017-053302] [Medline: 29437985]
52. Gilbert M, Thomson K, Salway T, Haag D, Grennan T, Fairley CK, et al. Differences in experiences of barriers to STI testing between clients of the internet-based diagnostic testing service GetCheckedOnline.com and an STI clinic in Vancouver, Canada. Sex Transm Infect 2019 Mar 07;95(2):151-156 [FREE Full text] [doi: 10.1136/sextrans-2017-053325] [Medline: 29437984]
53. Kuder M, Goheen MJ, Dize L, Barnes M, Gaydos C. Evaluation of a new website design for iwantthekit for chlamydia, gonorrhea, and trichomonas screening. Sex Transm Dis 2015 May;42(5):243-245 [FREE Full text] [doi: 10.1097/OLQ.0000000000000270] [Medline: 25868135]
54. Ling S, Richardson D, Mettenbrink C, Westergaard B, Sapp-Jones T, Crane L, et al. Evaluating a web-based test results system at an urban STI clinic. Sex Transm Dis 2010 Apr;37(4):259-263. [doi: 10.1097/OLQ.0b013e3181d3d037] [Medline: 20220562]
55. Mák G, Fowler H, Leaver C, Hagens S, Zelmer J. The effects of web-based patient access to laboratory results in British Columbia: a patient survey on comprehension and anxiety. J Med Internet Res 2015 Aug 04;17(8):e191 [FREE Full text] [doi: 10.2196/jmir.4350] [Medline: 26242801]
56. Brown L, Tan K, Guerra L, Naidoo C, Nardone A. Using behavioural insights to increase HIV self-sampling kit returns: a randomized controlled text message trial to improve England's HIV self-sampling service. HIV Med 2018 Oct 02;19(9):585-596 [FREE Full text] [doi: 10.1111/hiv.12634] [Medline: 29963766]
57. Kersaudy-Rahib D, Lydié N, Leroy C, March L, Bébéar C, Arwidson P, et al. Chlamyweb Study II: a randomised controlled trial (RCT) of an online offer of home-based sampling in France. Sex Transm Infect 2017 May 04;93(3):188-195. [doi: 10.1136/sextrans-2015-052510] [Medline: 28377422]
58. Reagan M, Xu H, Shih S, Secura G, Peipert J. A randomized trial of home versus clinic-based sexually transmitted disease screening among men. Sex Transm Dis 2012 Nov;39(11):842-847 [FREE Full text] [doi: 10.1097/OLQ.0b013e3182649165] [Medline: 23064532]
59. Wilson E, Free C, Morris TP, Syred J, Ahamed I, Menon-Johansson AS, et al. Internet-accessed sexually transmitted infection (e-STI) testing and results service: a randomised, single-blind, controlled trial. PLoS Med 2017 Dec 27;14(12):e1002479 [FREE Full text] [doi: 10.1371/journal.pmed.1002479] [Medline: 29281628]
60. Wilson E, Leyrat C, Baraitser P, Free C. Does internet-accessed STI (e-STI) testing increase testing uptake for chlamydia and other STIs among a young population who have never tested? Secondary analyses of data from a randomised controlled trial. Sex Transm Infect 2019 Dec 07;95(8):569-574 [FREE Full text] [doi: 10.1136/sextrans-2019-053992] [Medline: 31175210]
61. Ahmed-Little Y, Bothra V, Cordwell D, Powell D, Ellis D, Klapper P, et al. Attitudes towards HIV testing via home-sampling kits ordered online (RUClear pilots 2011-12). J Public Health (Oxf) 2016 Sep;38(3):585-590. [doi: 10.1093/pubmed/fdv075] [Medline: 26045472]
62. Witzel TC, Wright T, McCabe L, Gabriel MM, Wolton A, Gafos M, et al. Impact and acceptability of HIV self-testing for trans men and trans women: a mixed-methods subgroup analysis of the SELPHI randomised controlled trial and process evaluation in England and Wales. EClinicalMedicine 2021 Feb;32:100700 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100700] [Medline: 33681732]
63. Knight RE, Chabot C, Carson A, Thomson K, Haag D, Gilbert M, et al. Qualitative analysis of the experiences of gay, bisexual and other men who have sex with men who use: a comprehensive internet-based diagnostic service for HIV and other STIs. Sex Transm Infect 2019 Mar 13;95(2):145-150 [FREE Full text] [doi: 10.1136/sextrans-2018-053645] [Medline: 30636705]
64. Grandahl M, Larsson M, Herrmann B. 'To be on the safe side': a qualitative study regarding users' beliefs and experiences of internet-based self-sampling for and testing. BMJ Open 2020 Dec 29;10(12):e041340 [FREE Full text] [doi: 10.1136/bmjopen-2020-041340] [Medline: 33376171]
65. Robinson S, Reed M, Quevillon T, Hirvi E. Patient perceptions and interactions with their web portal-based laboratory results. BMJ Health Care Inform 2019 Apr 17;26(1) [FREE Full text] [doi: 10.1136/bmjhci-2019-000012] [Medline: 31039117]
66. Steehler K, Siegler AJ. Bringing HIV self-testing to scale in the United States: a review of challenges, potential solutions, and future opportunities. J Clin Microbiol 2019 Nov;57(11). [doi: 10.1128/jcm.00257-19]
67. Forrest S, McCaffery K, Waller J, Desai M, Szarewski A, Cadman L, et al. Attitudes to self-sampling for HPV among Indian, Pakistani, African-Caribbean and white British women in Manchester, UK. J Med Screen 2004 Jun 22;11(2):85-88 [FREE Full text] [doi: 10.1258/096914104774061065] [Medline: 15153323]
68. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
69. Sibbald B. Direct access to diagnostic services. Br J Gen Pract 2009 May 01;59(562):e144-e145. [doi: 10.3399/bjgp09x420563]

70. Balfe M, Brugha R. Disclosure of STI testing activities by young adults: the influence of emotions and social networks. Sociol Health Illn 2010 Nov;32(7):1041-1058 [FREE Full text] [doi: 10.1111/j.1467-9566.2010.01281.x] [Medline: 20937054]
71. Versluis A, van Luenen S, Meijer E, Honkoop PJ, Pinnock H, Mohr DC, et al. SERIES: eHealth in primary care. Part 4: addressing the challenges of implementation. Eur J Gen Pract 2020 Dec 07;26(1):140-145 [FREE Full text] [doi: 10.1080/13814788.2020.1826431] [Medline: 33025820]
72. Nagura-Ikeda M, Imai K, Tabata S, Miyoshi K, Murahara N, Mizuno T, et al. Clinical evaluation of self-collected saliva by Quantitative Reverse Transcription-PCR (RT-qPCR), Direct RT-qPCR, Reverse Transcription–Loop-Mediated Isothermal Amplification, and a Rapid Antigen Test to diagnose COVID-19. J Clin Microbiol 2020 Aug 24;58(9). [doi: 10.1128/jcm.01438-20]
73. Watkins I, Xie B. eHealth literacy interventions for older adults: a systematic review of the literature. J Med Internet Res 2014 Nov 10;16(11):e225 [FREE Full text] [doi: 10.2196/jmir.3318] [Medline: 25386719]
74. Spoelman WA, Bonten TN, de Waal MW, Drenthen T, Smeele IJ, Nielen MM, et al. Effect of an evidence-based website on healthcare usage: an interrupted time-series study. BMJ Open 2016 Nov 09;6(11):e013166 [FREE Full text] [doi: 10.1136/bmjopen-2016-013166] [Medline: 28186945]

Abbreviations

HCP: health care professional
MMAT: Mixed Method Appraisal Tool
MSM: men who have sex with men
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
STI: sexually transmitted infection

Edited by A Mavragani; submitted 01.04.21; peer-reviewed by J Hartvigsen, R McGowan; comments to author 14.07.21; revised version received 14.10.21; accepted 01.12.21; published 12.01.22.

Please cite as:
Versluis A, Schnoor K, Chavannes NH, Talboom-Kamp EPWA
Direct Access for Patients to Diagnostic Testing and Results Using eHealth: Systematic Review on eHealth and Diagnostics
J Med Internet Res 2022;24(1):e29303
URL: https://www.jmir.org/2022/1/e29303
doi:10.2196/29303
PMID:35019848

Review

Engagement Strategies to Improve Adherence and Retention in Web-Based Mindfulness Programs: Systematic Review

Natalie Winter[1], PhD; Lahiru Russell[1,2,3], PhD; Anna Ugalde[1], PhD; Victoria White[4], PhD; Patricia Livingston[1,5], PhD

[1]Centre for Quality and Patient Safety Research in the Institute for Health Transformation, Deakin University, Geelong, Australia

[2]School of Nursing and Midwifery, Deakin University, Geelong, Australia

[3]Centre for Quality and Patient Safety Research - Eastern Health Partnership, Deakin University, Melbourne, Australia

[4]School of Psychology, Faculty of Health, Deakin University, Geelong, Australia

[5]Faculty of Health, Deakin University, Geelong, Australia

Corresponding Author:
Natalie Winter, PhD
Centre for Quality and Patient Safety Research in the Institute for Health Transformation
Deakin University
1 Gheringhap Street
Geelong, 3220
Australia
Phone: 61 3 5227 1100
Email: n.heynsbergh@deakin.edu.au

Abstract

Background: Web-based mindfulness programs may be beneficial in improving the well-being outcomes of those living with chronic illnesses. Adherence to programs is a key indicator in improving outcomes; however, with the digitization of programs, it is necessary to enhance engagement and encourage people to return to digital health platforms. More information is needed on how engagement strategies have been used in web-based mindfulness programs to encourage adherence.

Objective: The aim of this study is to develop a list of engagement strategies for web-based mindfulness programs and evaluate the impact of engagement strategies on adherence.

Methods: A narrative systematic review was conducted across the MEDLINE Complete, CINAHL Complete, APA PsycINFO, and Embase databases and followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) guidelines. Articles were screened using the population, intervention, comparator, and outcome framework. Adults aged >18 years with chronic health conditions were included in the study. Mindfulness interventions, including those in combination with mindfulness-based cognitive therapy, delivered on the web through the internet or smartphone technology were included. Interventions lasted at least 2 weeks. Studies with a randomized controlled trial design or a pilot randomized controlled trial design were included. Engagement strategies, including web-based program features and facilitator-led strategies, adherence, and retention, were included.

Results: A total of 1265 articles were screened, of which 19 were relevant and were included in the review. On average, 70.98% (2258/3181) of the study participants were women with a mean age of 46 (SD 13) years. Most commonly, mindfulness programs were delivered to people living with mental health conditions (8/19, 42%). Of the 19 studies, 8 (42%) used only program features to encourage adherence, 5 (26%) used facilitator-led strategies, and 6 (32%) used a combination of the two. Encouraging program adherence was the most common engagement strategy used, which was used in 77% (10/13) of the facilitator-led studies and 57% (8/14) of the program feature studies. Nearly two-thirds (63%) of the studies provided a definition of adherence, which varied between 50% and 100% completion across studies. The overall mean participant compliance to the mindfulness programs was 56% (SD 15%). Most studies (10/19, 53%) had a long-term follow-up, with the most common follow-up period being 12 weeks after intervention (3/10, 30%). After the intervention, the mean retention was 78% (SD 15%).

Conclusions: Engagement strategies in web-based mindfulness programs comprise reminders to use the program. Other features may be suitable for encouraging adherence to interventions, and a facilitator-led component may result in higher retention. There is variance in the way adherence is measured, and intervention lengths and follow-up periods are inconsistent. More thorough reporting and a standardized framework for measuring adherence are needed to more accurately assess adherence and engagement strategies.

(J Med Internet Res 2022;24(1):e30026) doi:10.2196/30026

KEYWORDS

chronic disease; chronic illness; digital health; digital technology; internet mindfulness; mindfulness based stress reduction; patient dropouts; mobile phone

Introduction

Background

Mindfulness is the act of bringing awareness to the present moment in a nonjudgmental and accepting way [1]. Mindfulness programs are increasing in popularity as nonpharmacological alternatives to manage both physiological and psychological outcomes related to health conditions [2]. Psychological benefits are evident in individuals across a variety of conditions, including cancer [3] and mental illness [4], and physical health outcomes have been observed through improved blood pressure control [5] and improved glycemic control in people living with diabetes [6].

Evidence shows that mindfulness skills can be improved through greater engagement with meditation, home practice, face-to-face contact with a facilitator, and a higher number of sessions per week [7]. High adherence to both face-to-face and web-based mindfulness programs results in significant improvements in well-being outcome measures [8,9].

Mindfulness programs are increasingly being adapted to web-based platforms, providing opportunities for more people to participate compared with conventional face-to-face sessions [10]. Typically, adherence to web-based interventions is low, both with program adherence and study attrition [11]. Program adherence is poorly defined but needs to be standardized across studies; however, it is commonly conceptualized by the number of log-ins or number of sessions or modules completed in a program [12].

Adherence to web-based programs in previous reports has varied between 39.5% and 92% [9] compared with adherence to face-to-face settings, where the rates ranged between 26% and 100% [13] (based on definitions of 100% program completion). Mindfulness programs are often 8 weeks long in duration [9], with higher adherence having an impact on improved participant outcomes [14]. There is a need to explore whether engagement strategies can improve adherence to unmoderated web-based interventions. High attrition in telehealth interventions is common and can undermine the potential impact of programs [15]. Adherence to mindfulness-based interventions is often poorly defined and inconsistent across studies [16]. Promoting long-term adherence and engagement with web-based interventions may maximize the potential outcomes [17].

Engagement refers to the frequency and duration of use of the interventions, such as logging in and out of programs [18]. Strategies to support engagement are used to encourage and draw people back to the interventions [18]. Engagement can be enhanced by the design and features of web-based interventions, including the use of gamification, breaking content into manageable blocks, and using a variety of formats to deliver content such as video and visuals [19]. Other considerations for improving engagement include guided interaction from trained personnel [18], asynchronous emails [20], or web-based features such as reminders [18]. Behavior change techniques are engagement strategies incorporated into interventions to promote sustainable changes in behavior [21]. Behavior change techniques, such as notifications and semiautomated tracking, have previously been adopted in app-based interventions and have shown a positive impact on improving engagement [22]. In mindfulness programs, engagement involves regular meditation and daily awareness exercises coupled with intention motivation and commitment to practice [23]. Techniques such as self-reflection are incorporated into mindfulness programs and have been shown to positively impact symptoms in people with anxiety and stress [24]. More recent techniques such as machine learning [25] may also be used to tailor interventions to user-specific needs, thereby maximizing the clinical outcomes of users.

The influence of engagement strategies on program adherence has not been compared across studies; however, it is an important consideration when designing and implementing web-based interventions. In this review, we explored the engagement strategies applied in web-based mindfulness programs and evaluated whether these strategies had an impact on program adherence and retention.

Research Question

The following research question was used in the study: how can engagement strategies be incorporated into web-based mindfulness programs to improve adherence and retention?

Objectives

The objectives of this study are (1) to develop a list of engagement strategies for web-based mindfulness programs and (2) to evaluate the impact of engagement strategies on adherence.

Methods

Search Process

This systematic review was guided by the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) framework [26]. The following databases were searched for terms related to mindfulness, web-based programs, and engagement strategies: MEDLINE Complete, CINAHL Complete, APA PsycINFO, and Embase. The literature search focused on identifying papers published between January 2015 and March 2020. A 5-year period was chosen to capture the most recent web-based interventions. See Table S1 in Multimedia Appendix 1 for an example of the search strategy applied to the MEDLINE database. The reference lists of relevant articles and systematic reviews were searched for additional articles.

Eligibility Criteria

To guide the eligibility and screening process, the PICO (population, intervention, comparator, and outcome) framework [27] was used:

Population

Adults aged ≥18 years with a diagnosed chronic health condition or self-reported anxiety or depression were included in the study.

Intervention

Mindfulness interventions delivered on the web through the internet or smartphone technology were included. Mindfulness programs were defined as those focusing specifically on mindfulness-based practice, including programs using a combination of mindfulness and cognitive behavioral therapy (mindfulness-based cognitive therapy).

To allow for engagement strategies and adherence to be analyzed, the interventions had to be at least 2 weeks in duration. There is limited research to describe how long interventions should be to warrant the inclusion of engagement strategies. Previously, engagement was measured by reflecting on the previous 2 weeks [23]. Therefore, we determined that interventions had to be at least two weeks in duration to be included in the review.

Comparator and Context

Studies were required to have a comparison group with a randomized controlled trial (RCT) or a pilot RCT design.

Mindfulness programs developed by research groups for specific populations or commercially available mindfulness programs were tested in controlled trial settings.

Outcomes

Program adherence, study retention rate (%), and strategies such as web-based program features and facilitator-led features were included.

Screening

Retrieved articles were uploaded and managed by Endnote X9 (Clarivate Analytics). Duplicates were removed, and titles and abstracts were screened by 1 author (NH). Full-text articles were uploaded to *Covidence* to allow cross-checking between authors [28]. Full texts were reviewed independently by 2 authors, and any disagreements were resolved through discussion.

Data Extraction

A data extraction tool was developed in Microsoft Excel to standardize the extraction. Data were extracted by 1 author (NH), and 10% were cross-checked by the second author (PL).

Study Characteristics

Study data including author, year of publication, country, design, number of participants, intervention type, intervention duration, follow-up measurements, prior mindfulness experience, recruitment method, financial compensation, commercial app name, primary outcome, and primary findings were extracted.

Participant Characteristics

Gender, age, race, ethnicity, type of chronic illness or condition, and patient and caregiver status were extracted.

Adherence

Studies were included in the review when they reported per-protocol and intention-to-treat analyses. Because of variance in reporting the intervention, adherence was assessed in 3 different ways depending on the data available:

1. As a percentage of compliance with the intervention protocol. For example, some authors defined adherence as 80% program completion, and in this review, we recorded the percentage of the sample that was adherent with 80% program completion.
2. In groups defined by the study authors. For example, in an 8-week program, some authors reported the percentage of people who were adherent with 0- 3 sessions, 4- 6 sessions, and 7-8 sessions. In this review, we recorded the percentage of the sample that was adherent with the highest group of completion, for example, 7-8 sessions.
3. Summarized findings of the frequency and duration of use.

Retention

Retention rates were reported for the intervention group at postintervention measurements and subsequent follow-up points.

Engagement Strategies

Engagement strategies were categorized into following 3 groups:

1. Program features, including chat rooms, discussion boards, diaries and reflective processes, automated reminders, social support, goal setting, mood tracking, customization of content, demonstrations of meditation practice, and immediate feedback on meditation practice;
2. Facilitator-led strategies, including reminders from the research team to continue practice, contact with the research team to discuss practice or monitoring, and response to well-being scores throughout the intervention; and
3. A combination of program features and facilitator-led strategies.

Data Analysis

Study characteristics, participant characteristics, adherence, and retention rates were analyzed using descriptive statistics.

Data analysis consists of the following:

1. Exploring adherence: how adherence was defined, the impact of adherence on outcomes, impact of financial compensation on adherence, and impact of intervention length on adherence.
2. Describing retention at postintervention measurements and the last data collection point.
3. Describing engagement strategies (program features, facilitator-led strategies, or a combination): engagement strategies were categorized and summarized using frequency statistics.
4. Assessing the impact of engagement strategies on adherence: the relationship between engagement strategies and adherence was analyzed by comparing the type of

engagement strategy (program features, facilitator-led strategies, or a combination) with the percentage of people who reached program adherence or the percentage of people who adhered with the highest group of sessions (eg, those who completed 7-8 sessions in an 8-week program, as defined by the study authors).

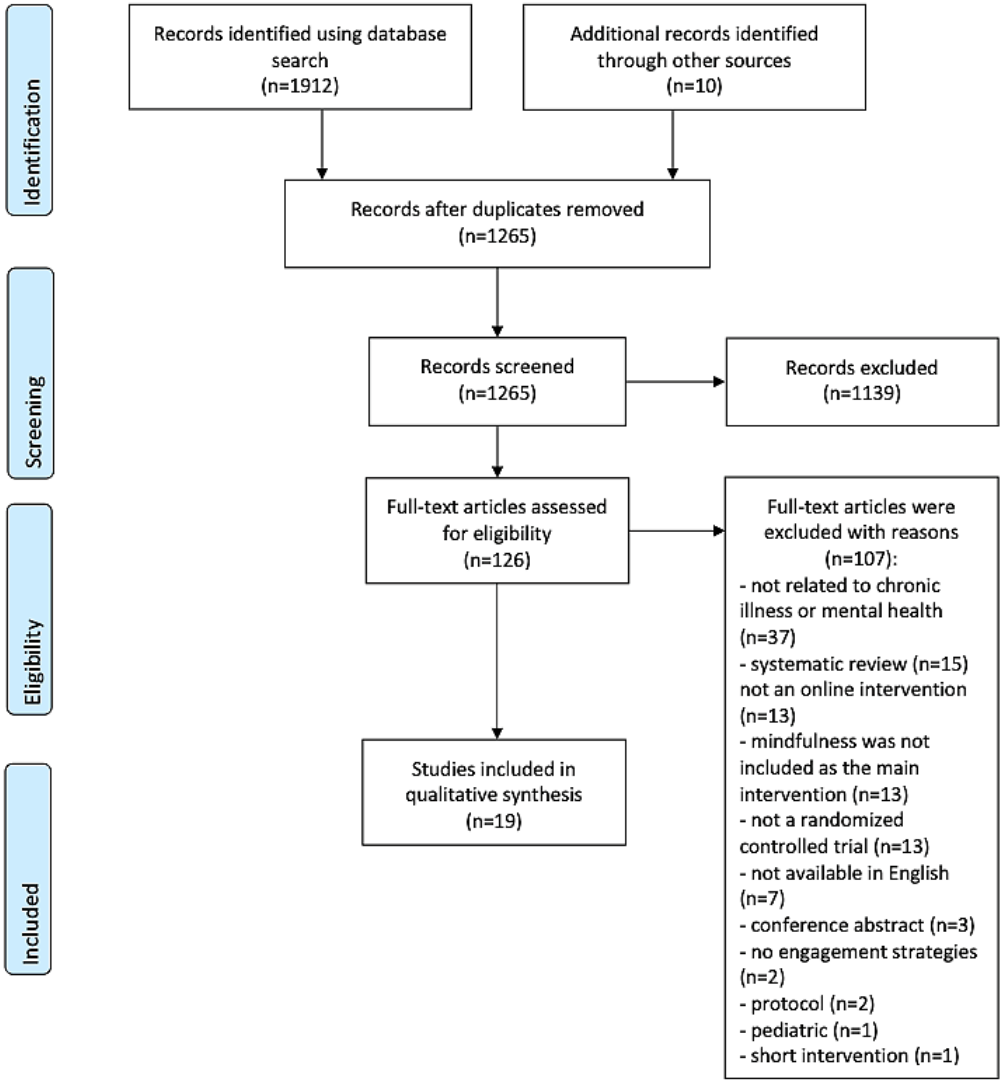
- Assessing the impact of engagement strategies on retention: the relationship between engagement strategies and retention was measured by comparing the type of engagement strategy with the intervention length, retention at the postintervention measurement and retention at the last follow-up points.

Results

Study Characteristics

A total of 1922 articles were retrieved from the databases and reference lists. After removing duplicates, a total of 1265 articles were screened by title and abstract. Full texts were retrieved for 126 articles, of which 19 were included in the review (Figure 1). Most studies were conducted in the United States (9/19, 47%) [5,24,29-35], were RCTs (16/19, 84%) [5,14,24,29,31-33,35-43], web-based (11/19, 58%) [3,14,33-35,37-43], and focused specifically on mindfulness or meditation (15/19, 79%) [3,5,14,29-32,34-36,39-43] (Table S2 in Multimedia Appendix 1) [3,5,14,24,27,29-41,43]. More studies (10/19, 52%) excluded people with previous or recent mindfulness experiences [24,29-31,37,39,42,43] than those who allowed participants with prior mindfulness experience (6/19, 31%) [3,14,32,34,35,38,40,41]. Over half of the studies used a combination of web-based and face-to-face recruitment strategies (10/19, 52%) [5,29,31,32,34,35,37,39,40,42]. Commercially available mindfulness apps, including Headspace (n=3) [30,32,36], Calm (n=1) [29], and Pacifica (n=1), were used by 5 (26%) studies [24]. A total of 3 (16%) studies provided monetary compensation for participation [29,30,33], and 3 (16%) provided access to paid mindfulness apps [30,32,36]. Intervention duration ranged from 2 weeks to 12 months, with over half (10/19, 53%) of the studies having an intervention duration of 8 weeks [29,30,32,33,36,37,39,40,42,43].

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) diagram of the search process.



A total of 8 (47%) studies focused on psychological measures as their primary outcome [29,31-34,37-39,42], and 3 (16%) used a physiological measure [5,14,41]. A total of 7 (37%) studies did not report the primary outcome [3,24,30,35,36,40,43]. Secondary outcomes were predominately psychological measures and program evaluations (8/19, 42%) [16,24,30,32,34,38,39,42]. Most (17/19, 90%) studies showed that mindfulness resulted in a significant improvement in outcomes either psychological or physical [3,5,14,24,29-34,36,37,39-43] (Table S3 in Multimedia Appendix 1) [3,5,14,27,29-41,43].

Participant Characteristics

A total of 34,601 participants were included in the trials. The mean sample size was 165 (SD 134; range 21-500). On average, 71% of the participants were women (SD 18; range 46-100) and 46 years old (SD 13; range 21-76). A total of 8 (42%) studies reported the ethnicity of the participants [24,29-35,39], and 4 (26%) reported race [29,31-33]. On average, White people comprised 74% (SD 14%) of the sample and 90% (SD 10%) were non-Hispanic. Mindfulness programs were delivered to people with a variety of chronic illnesses, with the most common conditions related to mental health (8/19, 42%) [24,29,31,33-36,38] and cancer (4/19, 21%) [3,30,32,37]. Most (17/21, 81%) studies were delivered to people living with the illness [3,5,14,24,29,31-41,43].

Engagement Strategies

A total of 8 (42%) studies used only program features to encourage adherence, 5 (26%) used only facilitator-led strategies, and 6 (32%) used a combination of the two (Table 1).

Within the facilitator-led strategies (n=13) [5,30,31,33-38,42,43], encouraging adherence was most commonly done using contact and reminders from facilitators to use the program (10/13, 77%) [30,31,33-36,38,42]. Contact with a facilitator for discussion of content or well-being scores was used to a lesser extent (4/13, 31%) [5,37,38,43]. In 7 (37%) studies, engagement with facilitators occurred weekly [34-37,42,43].

Table 1. Types of engagement strategies used across studies and their adherence rates.

Study	Program engagement	Facilitator engagement	Adherence with study protocol (%)
Chandler et al [5]	✓	✓	39
Compen et al [37]	✓	✓	79
Kladnitski et al [38]	✓	✓	66
Kubo et al [30]	✓	✓	56
Stjernsward and Hansson [27]	✓	✓	57
Thompson et al [33]	✓	✓	NR[a]
Gotink et al [14]	✓	N/A[b]	50
Hearn and Finlay [39]	✓	N/A	72
Henriksson et al [40]	✓	N/A	58
Huberty et al [29]	✓	N/A	NR
Moberg et al [24]	✓	N/A	NR
Rosen et al [32]	✓	N/A	NR
Russell et al [3]	✓	N/A	NR
Younge et al [41]	✓	N/A	53
Bostock et al [36]	N/A	✓	27
Lindsay et al [31]	N/A	✓	NR
Tavallaei et al [43]	N/A	✓	NR
Wahbeh et al [35]	N/A	✓	NR
Wahbeh [34]	N/A	✓	NR

[a]NR: not recorded.

[b]N/A: not applicable.

Within program feature strategies (n=14) [3,5,14,24,29,30,32,33,37-42], participants in 57% (8/14) studies received automated reminders [3,5,14,29,30,32,40,41]. Half of the program reminders were received at least once a week [3,14,32,41], and the remaining were sent on an ad hoc basis [29,30] or participants were able to personalize whether they received notifications or not [5,30]. Other program features used to encourage adherence included the ability to personalize mindfulness course content (4/14, 29%) [5,25,26,28], homework activities (3/14, 21%) [33,37,38], self-reflections (2/14, 14%) [37,42], social contact (3/14, 21%) [5,24,33], personalization of app appearance (2/14, 14%) [5,24], lesson summaries (1/14,

7%) [38], progress tracking of mindfulness practice (1/14, 7%)[24], immediate feedback on practice (1/14, 7%) [5], demonstration videos (1/14, 7%) [39], goal setting (1/14, 7%) [24], tracking psychological outcomes (1/14, 7%) [24], and tracking physical health (1/14, 7%) [24] (Table S4 in Multimedia Appendix 1) [3,5,14,24,29-35,37-41,43].

Contact initiated by facilitators or program reminders was most commonly delivered by email (9/14, 64%) [3,14,33,36-38,40-42] or telephone (7/14, 50%) [30,31,33-35,38,43].

Adherence

Nearly two-thirds (12/19, 63%) of the studies provided a definition of program adherence [3,5,14,30,33,36-42]. When defined as the percentage of program completion, the definitions of adherence varied between 50% and 100% program completion across studies. When adherence was grouped, the highest group of completion varied from 50% to 100% among the studies. A total of 6 (32%) studies did not provide a measurement for adherence and analyzed program use descriptively [24,29,31,32,34,35]. Moreover, 1 (5%) study did not report adherence or program use [43]. The percentage of people who complied with the authors' definitions of adherence ranged from 27% to 79%, with a mean compliance of 56% (SD 15%).

The Impact of Engagement Strategies on Adherence

Among studies that used only program features (n=8) [3,14,24,29,32,39-41], 4 recorded adherence between 50% and 72% (mean 58%, SD 8%) [14,39-41] (Table S3 in Multimedia Appendix 1). Among studies that used only facilitator-led strategies (n=5) [31,34-36,43], only 1 reported adherence of 27% [36]. Among studies that used a combination of program features and facilitator-led strategies (n=6) [5,30,33,37,38,42], 7 recorded adherence between 39% and 79% (mean 59%, SD 13) [5,30,37,38,42].

When examining studies that used program features, of the studies that used 1 strategy (n=6) [3,14,39-42], 5 measured adherence rates between 50% and 72% (mean 58%, SD 8%). Of the studies that used 2 strategies (n=8) [29,30,32,33,37,38], 3 measured adherence between 56% and 79% (mean 67%, SD 9%). A total of 5 studies did not include any engagement strategies within their program [15,34-36,43], and 2 [5,24] used ≥5 strategies; adherence was only recorded in 2 of these studies, and they were below 40%. Studies that involved only program reminders as engagement strategies (n=4, 3 studies recorded adherence) [3,14,40,41] had an average adherence rate of 54% (SD 3%) compared with the average adherence rate of 48% (SD 8%) of those studies that used reminders and other strategies (n=4, only 2 recorded adherence) [5,29,30,32], and the average adherence rate of 69% of those studies that did not use program reminders but only used other strategies (n=6, 4 recorded adherence) [24,33,37-39,42].

How Adherence Affected Outcomes

A total of 10 (53%) studies analyzed the relationship between outcome variables and adherence [14,24,30,32,34,36,38-40,42]. Of them, 4 studies found that people who had higher adherence to mindfulness programs had a significantly higher improvement in outcomes [30,36,40,42]; 1 study found that people with higher scores for depression at baseline were less likely to be adherent or complete mindfulness programs [39]; 1 found that people with higher blood pressure readings were more likely to be compliant [14]; and 1 showed that higher quality of life scores at baseline were significantly associated with improved adherence [32]. A total of 3 studies found no relationship between baseline scores and adherence or adherence and outcome variables [24,34,38].

Financial Compensation and Program Adherence

Of the 6 studies that provided any type of compensation, 2 measured adherence with a mean of 42% (SD 15%; range 27-56) [30,36]. Among the studies that did not offer financial compensation, the majority (8/13, 62%) measured adherence with a mean of 60% (SD 11%; range 39-79) [5,14,37-42].

Intervention Length and Program Adherence

The impact of the intervention length on adherence was analyzed. Of the 5 studies with an intervention <8 weeks, none recorded adherence. Those with an 8-week intervention recorded an average of 58% (SD 16%) adherence (6/10, 60% of the studies measured adherence) [30,36,37,39,40,42]; those with interventions >8 weeks recorded an average of 52% (SD 9%) adherence (4/4, 100% of the studies measured adherence) [5,14,38,41].

Retention

Most (10/19, 53%) studies conducted pre-post analysis with additional follow-up points [14,24,29,32-34,36,38,39,42]. Follow-up periods ranged from 4 to 36 weeks after intervention, and the most frequent follow-up time was 12 weeks after the intervention (3/10, 30%) [38,39,42]. After intervention, most (14/19, 74%) studies had over 70% retention (mean 78%, SD 15%; range 35%-100%) [3,5,29-31,33-39,41,43]. At the last follow-up point, 4 studies had retention above 70% [14,33,36,38].

The Impact of Engagement Strategies on Retention

Studies that applied only facilitator-led strategies, on average, were 6 weeks in duration (SD 2; range 2-8) and had a retention rate of 93% (SD 10; range 73-100) compared with studies with a combination of program features and facilitator-led strategies with a mean duration of 16 weeks (SD 10; range 8-52) and a retention rate of 75% (SD 5%; range 69-84) and those with only program features with a mean duration of 8 weeks (SD 2; range 4-12) and retention rate of 67% (SD 15%; range 30-79).

Of the studies that used facilitator-led strategies only, 40% (2/5) had follow-up periods after postintervention follow-up [34,36]. On average, follow-up was 7 (range 6-8) weeks and retention was 76% (SD 15%; range 69-82). Of the 6, 5 (50%) studies using a combination of program features and facilitator-led strategies had long-term follow-up, which, on average, was 11 weeks (SD 0.9; range 10-12), with a retention rate of 71% (SD 15; range 49-83) [33,38,42]. Of the 8, 5 (63%) studies using program features only also had a long-term follow-up period of, on average, 13 (range 4-36) weeks, with retention rates of 53% (SD 18; range 20-74) [14,24,29,32,39].

Studies that used only program reminders as engagement strategies (n=4) [3,14,40,41] had mean retention rates of 71% after intervention (n=3) [3,40,41] and retention of 74% at the last follow-up point (n=1) [14]. Studies that used reminders and other strategies (n=4) [5,29,30,32] had a mean retention of 78% after intervention (n=3) [5,29,30] and 57% at the last follow-up point (n=2) [29,32]. Studies that did not use program reminders but only used other strategies (n=8) [24,33,37-39,42] had a mean retention of 67% after intervention (n=6) [24,33,37-39,42] and 58% at the last follow-up point (n=5) [24,38,39,42].

Discussion

Principal Findings

In this review, we described the engagement strategies applied to web-based mindfulness programs and their impact on adherence rates. The use of program features only was associated with program adherence but not with maintaining study retention. Engagement strategies were largely reminders to use the program and, to a lesser extent, the ability to customize program content, interact with features, or engage with content on a deeper level through reflections, homework activities, and discussions of content with facilitators. There was little difference between the type of engagement strategy used and adherence to programs or retention rates.

The need to accurately report study and program attrition to better understand the associations between program adherence and health outcomes has been established [11,12]. Our review found variability across studies in adherence measurements and inconsistencies in reporting adherence. Some studies measured adherence as completing a specific percentage of the program [3,14,30,37,39,41]. Other studies described adherence by grouping the number of sessions completed [5,33,36,38,40,42] or by describing use [24,29,31,32,34,35]. Although findings suggest that program adherence is similar between interventions using program features only and those using a combination of program features and facilitator-led strategies, these results should be interpreted with caution because of the variability in reporting. The variability in measuring adherence is consistent in the e-therapy literature [44] and limits the ability to assess the relationships between adherence to and engagement with web-based interventions and user outcomes. Future studies should consider reporting adherence as a percentage of program completion for easier comparisons across studies.

Similarly, the ability to measure the impact of engagement strategies on study attrition is limited. The findings suggest that studies using only facilitator-led strategies were favorable for maintaining study retention [31,34-36,43]. On average, at the postintervention measurement, studies with only facilitator-led strategies had a retention rate of 93% (SD 10%) compared with the rate of those using only program features of 67% (SD 15%). Similar findings were observed during the follow-up period (76%, SD 7% vs 53%, SD 18%a). However, there is limited evidence as to whether the presence of the facilitator was the reason for this variability or whether other factors such as intervention length, follow-up length, or demographic characteristics of participants contributed to attrition. For example, 1 study that used only program features to improve engagement had low retention after intervention (35%) and at the 8-week follow-up (20%) [24]. No information was provided regarding the reason for these high attrition rates, making it difficult to determine the cause of these findings. The use of a facilitator or therapist to guide web-based psychological programs has been debated [45,46]. Studies of cognitive behavioral therapy interventions found that the presence of a therapist as a facilitator improved symptoms of depression compared with interventions with no facilitator [46]. However, improvements in anxiety symptoms were similar across studies [46], and no information was provided about whether the presence of a facilitator affected adherence. Improvements in patient outcomes may also be explained by the presence of comorbidities, including physical and mental illnesses, on which mindfulness may have a positive impact [47]. Therefore, participation in a mindfulness program targeting 1 disease may have additional benefits for other comorbid conditions. Furthermore, studies that used only facilitator-led strategies experienced, on average, a higher retention rate, which is similar to previous reviews that have described that self-directed interventions often require low levels of support from facilitators [16]. The use of facilitators to encourage adherence, or therapists to deliver content, needs to be weighed against the sustainability goals, cost of the program and length of the intervention during trials, and potential scaling after implementation.

Most studies in this review showed that web-based mindfulness resulted in improvements in either psychological or physiological outcome measures [3,5,14,24,29-34,36,37,39-43]. Two key findings from this review further highlighted the relationship between study retention and baseline functioning of participants, where those with poorer psychological well-being at baseline were more likely to drop out [39], and those with higher adherence were more likely to experience greater improvements in outcomes [30,36,40,42]. This is similar to previous findings where higher levels of worry and rumination at baseline resulted in disengagement from mindfulness-based interventions [23]. Stricter measurements of adherence are required in future studies to fully understand the role of adherence in the success of interventions.

Program features applied throughout studies to enhance engagement varied according to the type and number of features available to users. Furthermore, the number and type of features included had similar impacts on program adherence and study retention, suggesting that there may not be one superior feature to be included in programs. Features such as diaries, reminders, and social connectedness are commonly used in interventions as behavior change techniques [21], and web-based features have been shown to be successful in improving user outcomes in other e-interventions [48]. Within mindfulness, more specific reporting is needed to assess how often users engage with each type of feature to determine the relationship among engagement strategies, adherence, and outcomes.

Limitations

Across studies, there was a large variance in interventions and in reporting adherence. These factors made it difficult to draw any firm conclusions from the data.

The sample of the included studies was predominately White and female, which limits the generalizability of these findings to other population groups.

This review aims to describe the influence of engagement strategies on adherence and retention among people living with chronic illnesses or conditions. Other studies measuring adherence to mindfulness in the general population may have provided additional information on the impact of engagement strategies. However, there is a need to evaluate engagement and adherence to web-based interventions, specifically in people living with chronic illness. People with chronic illness may be more likely to experience depression and anxiety symptoms than those without a chronic illness [49]. Lower mental well-being can affect the use of and engagement with web-based interventions.

Furthermore, the primary outcome of the review was to assess adherence, retention, and engagement strategies rather than to draw conclusions about the effectiveness of interventions on patient outcomes. As a result, the risk of bias assessment was less relevant.

Conclusions

Engagement strategies in web-based mindfulness programs largely comprise reminders to use the program. The impact of other features such as personalization, self-reflection activities, and lesson summaries on adherence requires further investigation. There is variance in the way adherence is measured, and intervention lengths and follow-up periods are inconsistent. More thorough reporting and a standardized framework for measuring adherence are needed to more accurately assess adherence and engagement strategies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study demographics, examples of search strategy, and adherence, retention, and engagement strategies and outcomes.
[DOCX File , 53 KB - jmir_v24i1e30026_app1.docx]

References

1. Kabat-Zinn J. Full Catastrophe Living; Using the Wisdom of Your Body and Mind to Face Stress, Pain and Illness. London, UK: Piatkus; 1996:59-66.
2. Niazi AK, Niazi SK. Mindfulness-based stress reduction: a non-pharmacological approach for chronic illnesses. N Am J Med Sci 2011 Jan;3(1):20-23 [FREE Full text] [doi: 10.4297/najms.2011.320] [Medline: 22540058]
3. Russell L, Ugalde A, Orellana L, Milne D, Krishnasamy M, Chambers R, et al. A pilot randomised controlled trial of an online mindfulness-based program for people diagnosed with melanoma. Support Care Cancer 2019 Jul 30;27(7):2735-2746. [doi: 10.1007/s00520-018-4574-6] [Medline: 30506103]
4. Potes A, Souza G, Nikolitch K, Penheiro R, Moussa Y, Jarvis E, et al. Mindfulness in severe and persistent mental illness: a systematic review. Int J Psychiatry Clin Pract 2018 Nov 07;22(4):253-261. [doi: 10.1080/13651501.2018.1433857] [Medline: 29411670]
5. Chandler J, Sox L, Diaz V, Kellam K, Neely A, Nemeth L, et al. Impact of 12-month smartphone breathing meditation program upon systolic blood pressure among non-medicated stage 1 hypertensive adults. Int J Environ Res Public Health 2020 Mar 17;17(6):1955 [FREE Full text] [doi: 10.3390/ijerph17061955] [Medline: 32192020]
6. Rosenzweig S, Reibel DK, Greeson JM, Edman JS, Jasser SA, McMearty KD, et al. Mindfulness-based stress reduction is associated with improved glycemic control in type 2 diabetes mellitus: a pilot study. Altern Ther Health Med 2007;13(5):36-38. [Medline: 17900040]
7. Strohmaier S. The relationship between doses of mindfulness-based programs and depression, anxiety, stress, and mindfulness: a dose-response meta-regression of randomized controlled trials. Mindfulness 2020 Mar 02;11(6):1315-1335 [FREE Full text] [doi: 10.1007/s12671-020-01319-4]
8. Ribeiro L, Atchley RM, Oken BS. Adherence to practice of mindfulness in novice meditators: practices chosen, amount of time practiced, and long-term effects following a mindfulness-based intervention. Mindfulness (N Y) 2018 Apr;9(2):401-411 [FREE Full text] [doi: 10.1007/s12671-017-0781-3] [Medline: 30881517]
9. Spijkerman M, Pots W, Bohlmeijer E. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
10. Mrazek AJ, Mrazek MD, Cherolini CM, Cloughesy JN, Cynman DJ, Gougis LJ, et al. The future of mindfulness training is digital, and the future is now. Curr Opin Psychol 2019 Aug;28:81-86. [doi: 10.1016/j.copsyc.2018.11.012] [Medline: 30529975]
11. Eysenbach G. The law of attrition. J Med Internet Res 2005;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]

12. Sieverink F, Kelders SM, van Gemert-Pijnen JE. Clarifying the concept of adherence to eHealth technology: systematic review on when usage becomes adherence. J Med Internet Res 2017 Dec 06;19(12):e402 [FREE Full text] [doi: 10.2196/jmir.8578] [Medline: 29212630]
13. Stanic J, Barth J, Danon N, Bondolfi G, Jermann F, Eicher M. Adherence to standardized 8-week mindfulness-based interventions among women with breast or gynecological cancer: a scoping review. J Psychosos Oncol Res Pract 2021;3(2):e048. [doi: 10.1097/OR9.0000000000000048]
14. Gotink RA, Younge JO, Wery MF, Utens EM, Michels M, Rizopoulos D, et al. Online mindfulness as a promising method to improve exercise capacity in heart disease: 12-month follow-up of a randomized controlled trial. PLoS One 2017;12(5):e0175923 [FREE Full text] [doi: 10.1371/journal.pone.0175923] [Medline: 28486559]
15. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
16. Russell L, Ugalde A, Milne D, Austin D, Livingston PM. Digital characteristics and dissemination indicators to optimize delivery of internet-supported mindfulness-based interventions for people with a chronic condition: systematic review. JMIR Ment Health 2018 Aug 21;5(3):e53 [FREE Full text] [doi: 10.2196/mental.9645] [Medline: 30131317]
17. Beintner I, Vollert B, Zarski A, Bolinski F, Musiat P, Görlich D, et al. Adherence reporting in randomized controlled trials examining manualized multisession online interventions: systematic review of practices and proposal for reporting standards. J Med Internet Res 2019 Aug 15;21(8):e14181 [FREE Full text] [doi: 10.2196/14181] [Medline: 31414664]
18. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
19. Ubhi HK, Michie S, Kotz D, van Schayck OC, Selladurai A, West R. Characterising smoking cessation smartphone applications in terms of behaviour change techniques, engagement and ease-of-use features. Transl Behav Med 2016 Sep 23;6(3):410-417 [FREE Full text] [doi: 10.1007/s13142-015-0352-x] [Medline: 27528530]
20. Arnold C, Villagonzalo K, Meyer D, Farhall J, Foley F, Kyrios M, et al. Predicting engagement with an online psychosocial intervention for psychosis: exploring individual- and intervention-level predictors. Internet Interv 2019 Dec;18:100266 [FREE Full text] [doi: 10.1016/j.invent.2019.100266] [Medline: 31890619]
21. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
22. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
23. Banerjee M, Cavanagh K, Strauss C. Barriers to mindfulness: a path analytic model exploring the role of rumination and worry in predicting psychological and physical engagement in an online mindfulness-based intervention. Mindfulness (N Y) 2018 Nov 6;9(3):980-992 [FREE Full text] [doi: 10.1007/s12671-017-0837-4] [Medline: 29875884]
24. Moberg C, Niles A, Beermann D. Guided self-help works: randomized waitlist controlled trial of Pacifica, a mobile app integrating cognitive behavioral therapy and mindfulness for stress, anxiety, and depression. J Med Internet Res 2019 Jun 08;21(6):e12556 [FREE Full text] [doi: 10.2196/12556] [Medline: 31199319]
25. Chien I, Enrique A, Palacios J, Regan T, Keegan D, Carter D, et al. A machine learning approach to understanding patterns of engagement with internet-delivered mental health interventions. JAMA Netw Open 2020 Jul 01;3(7):e2010791 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.10791] [Medline: 32678450]
26. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. J Clin Epidemiol 2009 Oct;62(10):1006-1012. [doi: 10.1016/j.jclinepi.2009.06.005] [Medline: 19631508]
27. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. BMC Med Inform Decis Mak 2007 Jun 15;7(1):16 [FREE Full text] [doi: 10.1186/1472-6947-7-16] [Medline: 17573961]
28. Covidence. Veritas Health Innovation Ltd, Australia. URL: https://www.covidence.org/ [accessed 2020-07-08]
29. Huberty J, Green J, Glissmann C, Larkey L, Puzia M, Lee C. Efficacy of the mindfulness meditation mobile app "Calm" to reduce stress among college students: randomized controlled trial. JMIR Mhealth Uhealth 2019 Jun 25;7(6):e14273 [FREE Full text] [doi: 10.2196/14273] [Medline: 31237569]
30. Kubo A, Kurtovich E, McGinnis M, Aghaee S, Altschuler A, Quesenberry C, et al. A randomized controlled trial of mHealth mindfulness intervention for cancer patients and informal cancer caregivers: a feasibility study within an integrated health care delivery system. Integr Cancer Ther 2019;18:1534735419850634 [FREE Full text] [doi: 10.1177/1534735419850634] [Medline: 31092044]
31. Lindsay EK, Young S, Smyth JM, Brown KW, Creswell JD. Acceptance lowers stress reactivity: dismantling mindfulness training in a randomized controlled trial. Psychoneuroendocrinology 2018 Jan;87:63-73. [doi: 10.1016/j.psyneuen.2017.09.015] [Medline: 29040891]

32. Rosen KD, Paniagua SM, Kazanis W, Jones S, Potter JS. Quality of life among women diagnosed with breast Cancer: a randomized waitlist controlled trial of commercially available mobile app-delivered mindfulness training. Psychooncology 2018 Aug 01;27(8):2023-2030. [doi: 10.1002/pon.4764] [Medline: 29766596]
33. Thompson NJ, Patel AH, Selwa LM, Stoll SC, Begley CE, Johnson EK, et al. Expanding the efficacy of Project UPLIFT: distance delivery of mindfulness-based depression prevention to people with epilepsy. J Consult Clin Psychol 2015 Apr;83(2):304-313 [FREE Full text] [doi: 10.1037/a0038404] [Medline: 25495361]
34. Wahbeh H. Internet Mindfulness Meditation Intervention (IMMI) improves depression symptoms in older adults. Medicines (Basel) 2018 Nov 02;5(4):119 [FREE Full text] [doi: 10.3390/medicines5040119] [Medline: 30400211]
35. Wahbeh H, Goodrich E, Oken B. Internet-based mindfulness meditation for cognition and mood in older adults: a pilot study. Altern Ther Health Med 2016;22(2):44-53 [FREE Full text] [Medline: 27036056]
36. Bostock S, Crosswell AD, Prather AA, Steptoe A. Mindfulness on-the-go: effects of a mindfulness meditation app on work stress and well-being. J Occup Health Psychol 2019 Feb;24(1):127-138 [FREE Full text] [doi: 10.1037/ocp0000118] [Medline: 29723001]
37. Compen F, Bisseling E, Schellekens M, Donders R, Carlson L, van der Lee M, et al. Face-to-face and internet-based mindfulness-based cognitive therapy compared with treatment as usual in reducing psychological distress in patients with cancer: a multicenter randomized controlled trial. J Clin Oncol 2018 Aug 10;36(23):2413-2421. [doi: 10.1200/jco.2017.76.5669]
38. Kladnitski N, Smith J, Uppal S, James MA, Allen AR, Andrews G, et al. Transdiagnostic internet-delivered CBT and mindfulness-based treatment for depression and anxiety: a randomised controlled trial. Internet Interv 2020 Apr;20:100310 [FREE Full text] [doi: 10.1016/j.invent.2020.100310] [Medline: 32140427]
39. Hearn JH, Finlay KA. Internet-delivered mindfulness for people with depression and chronic pain following spinal cord injury: a randomized, controlled feasibility trial. Spinal Cord 2018 Aug;56(8):750-761. [doi: 10.1038/s41393-018-0090-2] [Medline: 29581519]
40. Henriksson J, Wasara E, Rönnlund M. Effects of eight-week-web-based mindfulness training on pain intensity, pain acceptance, and life satisfaction in individuals with chronic pain. Psychol Rep 2016 Dec;119(3):586-607. [doi: 10.1177/0033294116675086] [Medline: 27780873]
41. Younge JO, Wery MF, Gotink RA, Utens EM, Michels M, Rizopoulos D, et al. Web-based mindfulness intervention in heart disease: a randomized controlled trial. PLoS One 2015;10(12):e0143843 [FREE Full text] [doi: 10.1371/journal.pone.0143843] [Medline: 26641099]
42. Stjernswärd S, Hansson L. Effectiveness and usability of a web-based mindfulness intervention for caregivers of people with mental or somatic illness. A randomized controlled trial. Internet Interv 2018 Jun;12:46-56 [FREE Full text] [doi: 10.1016/j.invent.2018.03.004] [Medline: 30135768]
43. Tavallaei V, Rezapour-Mirsaleh Y, Rezaiemaram P, Saadat SH. Mindfulness for female outpatients with chronic primary headaches: an internet-based bibliotherapy. Eur J Transl Myol 2018 Apr 24;28(2):7380 [FREE Full text] [doi: 10.4081/ejtm.2018.7380] [Medline: 29991985]
44. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
45. Fish J, Brimson J, Lynch S. Mindfulness interventions delivered by technology without facilitator involvement: what research exists and what are the clinical outcomes? Mindfulness (N Y) 2016 Jun;7(5):1011-1023 [FREE Full text] [doi: 10.1007/s12671-016-0548-2] [Medline: 27642370]
46. Saddichha S, Al-Desouki M, Lamia A, Linden IA, Krausz M. Online interventions for depression and anxiety - a systematic review. Health Psychol Behav Med 2014 Jan 1;2(1):841-881 [FREE Full text] [doi: 10.1080/21642850.2014.945934] [Medline: 25750823]
47. Alsubaie M, Abbott R, Dunn B, Dickens C, Keil TF, Henley W, et al. Mechanisms of action in mindfulness-based cognitive therapy (MBCT) and mindfulness-based stress reduction (MBSR) in people with physical and/or psychological conditions: a systematic review. Clin Psychol Rev 2017 Jul;55:74-91. [doi: 10.1016/j.cpr.2017.04.008] [Medline: 28501707]
48. Riva S, Camerini A, Allam A, Schulz PJ. Interactive sections of an internet-based intervention increase empowerment of chronic back pain patients: randomized controlled trial. J Med Internet Res 2014;16(8):e180 [FREE Full text] [doi: 10.2196/jmir.3474] [Medline: 25119374]
49. Parks AC, Williams AL, Kackloudis GM, Stafford JL, Boucher EM, Honomichl RD. The effects of a digital well-being intervention on patients with chronic conditions: observational study. J Med Internet Res 2020 Jan 10;22(1):e16211 [FREE Full text] [doi: 10.2196/16211] [Medline: 31922491]

Abbreviations

RCT: randomized controlled trial

Edited by R Kukafka; submitted 29.04.21; peer-reviewed by H Wahbeh, H Haskelberg; comments to author 18.06.21; revised version received 09.08.21; accepted 12.09.21; published 12.01.22.

Please cite as:
Winter N, Russell L, Ugalde A, White V, Livingston P
Engagement Strategies to Improve Adherence and Retention in Web-Based Mindfulness Programs: Systematic Review
J Med Internet Res 2022;24(1):e30026
URL: https://www.jmir.org/2022/1/e30026
doi:10.2196/30026
PMID:35019851

Review

Perceptions and Needs of Artificial Intelligence in Health Care to Increase Adoption: Scoping Review

Han Shi Jocelyn Chew[1], DPhil; Palakorn Achananuparp[2], DPhil

[1]Alice Lee Centre for Nursing Studies, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore

[2]Living Analytics Research Centre, Singapore Management University, Singapore, Singapore

Corresponding Author:
Han Shi Jocelyn Chew, DPhil
Alice Lee Centre for Nursing Studies
Yong Loo Lin School of Medicine
National University of Singapore
Level 3, Clinical Research Centre
Block MD11, 10 Medical Drive
Singapore, 117597
Singapore
Phone: 65 65168687
Email: jocelyn.chew.hs@nus.edu.sg

Abstract

Background: Artificial intelligence (AI) has the potential to improve the efficiency and effectiveness of health care service delivery. However, the perceptions and needs of such systems remain elusive, hindering efforts to promote AI adoption in health care.

Objective: This study aims to provide an overview of the perceptions and needs of AI to increase its adoption in health care.

Methods: A systematic scoping review was conducted according to the 5-stage framework by Arksey and O'Malley. Articles that described the perceptions and needs of AI in health care were searched across nine databases: ACM Library, CINAHL, Cochrane Central, Embase, IEEE Xplore, PsycINFO, PubMed, Scopus, and Web of Science for studies that were published from inception until June 21, 2021. Articles that were not specific to AI, not research studies, and not written in English were omitted.

Results: Of the 3666 articles retrieved, 26 (0.71%) were eligible and included in this review. The mean age of the participants ranged from 30 to 72.6 years, the proportion of men ranged from 0% to 73.4%, and the sample sizes for primary studies ranged from 11 to 2780. The perceptions and needs of various populations in the use of AI were identified for general, primary, and community health care; chronic diseases self-management and self-diagnosis; mental health; and diagnostic procedures. The use of AI was perceived to be positive because of its availability, ease of use, and potential to improve efficiency and reduce the cost of health care service delivery. However, concerns were raised regarding the lack of trust in data privacy, patient safety, technological maturity, and the possibility of full automation. Suggestions for improving the adoption of AI in health care were highlighted: enhancing personalization and customizability; enhancing empathy and personification of AI-enabled chatbots and avatars; enhancing user experience, design, and interconnectedness with other devices; and educating the public on AI capabilities. Several corresponding mitigation strategies were also identified in this study.

Conclusions: The perceptions and needs of AI in its use in health care are crucial in improving its adoption by various stakeholders. Future studies and implementations should consider the points highlighted in this study to enhance the acceptability and adoption of AI in health care. This would facilitate an increase in the effectiveness and efficiency of health care service delivery to improve patient outcomes and satisfaction.

(J Med Internet Res 2022;24(1):e32939) doi:10.2196/32939

KEYWORDS

artificial intelligence; health care; service delivery; perceptions; needs; scoping; review

Introduction

Background

Rapid advances in artificial intelligence (AI)—software systems designed to mimic human intelligence or cognitive functions—have sparked confidence in its potential to enhance the efficiency of health care service delivery and patient outcomes [1-3]. However, although AI has been rapidly adopted in many industries, such as finance and information technology (IT), its adoption in health care remains relatively lagged because of the ethical and safety considerations that are more pronounced when it comes to human lives at stake [4]. AI-powered systems in health care can autonomously or semiautonomously perform a wide variety of tasks, such as medical diagnosis [5], treatment [6], and self-monitoring and coaching [7,8]. In some studies, AI has been shown to outperform human capabilities, such as analyses of chest x-ray images by radiologists [9]. Not only is AI expected to improve the quality of care and health outcomes for patients by decreasing human errors, but it is also likely to free up time for clinicians and health care workers from routine and repetitive tasks, enabling them to focus on more complex tasks [9,10]. For instance, in many areas of medical imaging, the use of fast and accurate AI-assisted diagnoses would significantly increase the workflow efficiency by processing more than 250 million images per day [11]. Various AI chatbots have also been developed to provide mental health counseling and assist overburdened clinicians [9]. Through AI-enabled apps and wearable devices, patients and the public could self-monitor and self-diagnose symptoms, such as atrial fibrillation, skin lesions, and retinal diseases [9].

Owing to the emerging nature of modern AI systems, the perceptions and needs of affected stakeholders (eg, health care providers, patients, caregivers, policy makers, and IT technicians) on the use of AI in health care are not yet fully understood. A large body of literature suggests that human factors, such as trust, perceived usefulness, and privacy, play an important role in the acceptance and adoption of past technologies in health care, including handheld devices [12], IT [13], and assistive technologies [14]. However, current evidence remains broad and general, and little is known about the perceptions and needs of AI in community health care. As the world makes a paradigm shift from curative to preventive medicine, AI holds a strong transformative potential to enhance sustainable health care by empowering self-care, such as self-monitoring and self-diagnosis. However, it is important to first understand the perspectives of all direct users of AI-driven systems (eg, patients and frontline health workers) and their perceived needs to ensure its successful adoption across different parts of the health care sector, especially community health care. Thus, this study aims to present an overview of the perceptions and needs of AI in community health care. The implications of this study will help inform the design of future health care–related AI technology to better fit the needs of users and enhance the adoption and acceptability of the technology.

Definition of AI

First, as the term *AI* is broadly used in many disciplines to represent various forms of intelligent systems and algorithms, it is important to establish a concrete and unified definition of AI for this study. Specifically, we adopted the definition of AI proposed by the High-Level Expert Group on Artificial Intelligence [15], which describes AI in terms of both a technology and field of study:

> *Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.*
>
> *As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics which includes control, perception, sensors, and actuators, as well as the integration of all other techniques into cyber-physical systems.*

Furthermore, most, if not all, modern AI systems are considered artificial narrow intelligence (ANI) or *Weak AI* [15] designed to perform one or more specific tasks. In health care, domain-specific tasks for ANI may vary from performing human perceptions, such as image recognition [16] and natural language processing [17], to making complex clinical decisions, such as medical diagnostics [18]. Many recent advances and breakthroughs in ANI use learning-based approaches, namely, deep learning, in which computational models consisting of several layers of artificial neural networks (hence the titular *deep*) are trained by learning from a massive amount of sample data to perform specific tasks. Although recent performances of ANI appear very promising, ANI models are limited in their generalizability, that is, models trained to perform tasks in one domain cannot be generalized to other domains. For example, ANI trained to diagnose diabetic retinopathy from fundus images cannot be directly used to detect pneumonia from chest x-ray images. In contrast to ANI, artificial general intelligence (AGI) or *Strong AI* [15] belongs to a class of AI that displays true human intelligence, capable of continuously learning and performing any tasks like a real human. AGI is most likely in public consciousness when talking about AI, as it is frequently portrayed in popular culture by sentient robots and self-aware systems. At present, no AI systems have been able to come close to exhibit the AGI capability. For a useful and concise summary regarding the definitions, terminologies, and history of AI, see the following technical reports: Ethics Guidelines for

Trustworthy AI [15] and Historical Evolution of Artificial Intelligence [19].

Methods

A systematic scoping review was conducted according to the 5-stage framework by Arksey and O'Malley [20]. Results were reported according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist (Multimedia Appendix 1) [21].

Stage 1: Identifying the Research Question

Our research question was as follows: What is known about the perceptions and needs of AI in health care?

Stage 2: Identifying Relevant Studies

Studies were searched from inception until June 21, 2021, using a 3-step search strategy. First, potential keywords and Medical Subject Headings terms were generated through iterative searches on PubMed and Embase. Keywords such as machine learning did not result in better search outcomes (ie, many irrelevant results were retrieved, such as the use of machine learning to explore perceptions of other topics); hence, they were omitted. Next, keywords including *artificial intelligence*, *AI*; *public*; *consumer*; *community*; *perception**; *preference**; *needs**; *opinions**; and *acceptability* were searched through nine databases: ACM Library, CINAHL, Cochrane Central, Embase, IEEE Xplore, PsycINFO, PubMed, Scopus, and Web of Science. Additional articles were also retrieved from the first 10 pages of the Google Scholar search results and the reference lists of the included full-text articles. The specific database searches combined with Boolean operators are detailed in Multimedia Appendix 2.

Stage 3: Study Selection

After removing duplicate articles, titles and abstracts were first screened by HSJC for inclusion eligibility. Articles were included if they were (1) focused on the use of AI in health care, except those focused on using AI to improve surgical techniques; (2) focused on perceptions, needs, and acceptability of AI in health care; (3) empirical studies or systematic reviews; (4) on adults aged ≥18 years; and (5) used in a community setting. Articles were excluded if they were (1) not specific to AI (eg, general eHealth or mobile health); (2) pilot studies, commentaries, perspectives, or opinion papers; and (3) not presented in the English language. In total, 43 full-text articles were screened independently by both coauthors, and discrepancies were resolved through discussions and consensus.

Stage 4: Charting the Data

Data were extracted by HSJC using Microsoft Excel according to the following headings: author, year, title, aim, type of publication, study design, country, AI applications in health care, data collection method, population characteristics, sample size, age (mean or range), proportion of men, acceptability, perceptions, needs and preferences, and limitations.

Results

Stage 5: Collating, Summarizing, and Reporting Results

A total of 3666 articles were retrieved from the initial search. After removing duplicate articles, 50.74% (1860/3666) of titles and abstracts were screened, and 0.91% (17/1860) of full-text articles were excluded for reasons shown in Figure 1. A total of 1.4% (26/1860) of articles were included in this study, with the study characteristics summarized in Table 1 and detailed in Multimedia Appendix 3 [22-47]. The mean age of participants ranged from 30 to 72.6 years, and the proportion of men ranged from 0% to 73.4%. Sample sizes for studies with human subject responses ranged from 11 to 2780, and secondary data (ie, journal articles and app reviews) ranged from 31 to 1826 [22-24]. Interestingly, 19% (5/26) of studies focused on the use of chatbots in health care [23-27] and 31% (8/26) of studies measured acceptability using questionnaires, surveys, interviews [25,26,28-33], and the discrete choice experiment (Multimedia Appendix 4 [22-32,34,36,37,39,41-44,47]) [34]. All the studies showed at least moderate acceptability, or >50% of the participants showed acceptance toward the use of AI in health care, albeit only for minor conditions [26]. Age, IT skills, preference for talking to computers, perceived utility, positive attitude, and perceived trustworthiness were found to be associated with AI acceptability [25,26].

Figure 1. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flow diagram of search strategy. AI: artificial intelligence.

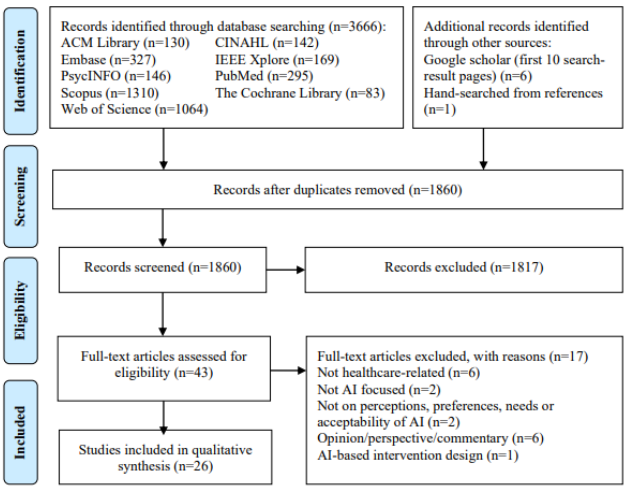


Table 1. Summary of study characteristics (N=26).

Study characteristics	Value, n (%)
Country	
Australia and New Zealand [35]	1 (4)
Canada [27,36-38]	4 (15)
China [22,32,33,39,40]	6 (23)
France [41]	1 (4)
India [24,42]	2 (8)
Korea [48]	1 (4)
Saudi Arabia [29]	1 (4)
Switzerland [30]	1 (4)
United Kingdom [23,26,31,43,44]	5 (19)
United Kingdom, Cyprus, Australia, the Netherlands, Sweden, Spain, United States, and Canada [28]	1 (4)
United States [25,45,46]	3 (12)
Type of publication	
Journal papers [22-29,31-41,43-47]	24 (92)
Conference papers [30,42]	2 (8)
Study design	
Observational [22,24,27-30,33-35,39,43-47]	15 (58)
Qualitative [36-38,41,42]	5 (19)
Mixed methods [25,26,31,32,40]	5 (19)
Systematic review [23]	1 (4)
Population characteristics	
General public [22,24,26,30,32-34,37,45]	9 (35)
Health care, government, technology, and industrial staff [27-29,35,36,40-44]	10 (39)
Patients and caregivers with specific diseases [25,31,34,36,38,39,47]	7 (27)
Mixture (systematic review) [23]	1 (4)
Artificial intelligence applications in health care	
General health care [22,23,26,27,29,33,36,37,40,41,43]	11 (42)
Primary [44] and community health care [28,42]	3 (12)
Chronic disease self-management [25,31,47]	3 (12)
Self-diagnosis [30,32,34,39]	4 (15)
Mental health [24,38]	2 (8)
Diagnostics [35,45,46]	3 (12)

Positive Perceptions

Overview

Several positive perceptions on the use of AI in health care were highlighted in our findings (Table 2).

Table 2. Perceptions on the use of artificial intelligence (AI) in health care.

Study	Available on demand and user-friendly	Efficiency	Price	Lack of trust in data privacy	Lack of trust in patient safety	Lack of trust in technology	Concerns over full automation
Abdi et al [28]	Able to collect data nonintrusively	Could support the self-care needs of older people—mobility, self-care and domestic life, social life and relationships, psychological support, and access to health care; potential uses for remote monitoring and prompting daily reminders, for example, medications	Cost was seen as both a facilitator of and a barrier to the older people's adoption of AI[a]	Especially in voice-activated devices	Deemed technically and commercially ready to support the care needs of older people	NS[b]	NS
Abdullah and Fakieh [29]	NS	Speeds up health care processes	NS	NS	AI was unable to provide opinions in unexpected situations	NS	Most health care employees feared that the AI would replace their job (mean score 3.11 of 4)
Baldauf et al [30]	Constant availability, not restricted by physical location	Quicker diagnosis and no waiting time	AI could be a cost-saving alternative	There were concerns over data privacy	Users were unsure about the legality of official medical certification and app trustworthiness	NS	Only a minority would rely solely on an AI-driven app for assessing health
Castagno and Khalifa [43]	NS	In all, 79% of health care staff believed AI could be useful or extremely useful in their field of work	NS	In all, 80% of health care staff believed there may be serious privacy issues	NS	NS	Overall, 10% of health care staff worried AI will replace their job
Easton et al [31]	NS	NS	NS	Patients were not concerned over data sharing	Patients were unsure whether to treat a chatbot as a real physician or an adviser	NS	NS
Gao et al [22]	NS	NS	NS	Distrust of AI companies accounted for a quarter of all negative opinions among social media users	Social media users were pessimistic about the immaturity of AI technology	NS	Less than half of the social media posts expressed that AI would completely or partially replace human doctors
Griffin et al [25]	NS	The majority were interested in using a chatbot to help manage medications, refills, communicate with care teams, and accountability toward self-care tasks	NS	There were concerns with chatbots providing too much information and invading privacy	There were concerns with chatbots making overwhelming demands for lifestyle changes	NS	NS
Kim [47]	NS	NS	NS	NS	NS	NS	NS

Study	Available on demand and user-friendly	Efficiency	Price	Lack of trust in data privacy	Lack of trust in patient safety	Lack of trust in technology	Concerns over full automation
Lai et al [41]	NS	NS	NS	There were legal difficulties to access individual health data; regulate use; strike balance between health, social justice, and freedom; and need to achieve confidentiality and respect for privacy	NS	NS	NS
Li et al [32]	NS	NS	NS	NS	AI may not understand complex emotional problems and give incurable diagnoses; and unsure whether doctors would accept the information provided by the AI	NS	NS
Liu et al [34]	NS	NS	NS	NS	Majority were confident that AI diagnosis methods would outperform human clinician diagnosis methods because of higher accuracy	NS	Majority preferred to receive combined diagnoses from both AI and human clinicians
Liu et al [39]	NS	NS	Acceptability depends on the expense of AI diagnosis compared with that of physicians	NS	Accuracy was deemed the most important attribute for AI uptake	NS	NS
Liyanage et al [44]	NS	Improves efficiency through decision support to improve primary health care processes and pattern recognition in imaging	NS	NS	There were concerns over the risk of medical errors, bias, and secondary effects of using AI (eg, insurance)	NS	AI technology is still not competent to replace human decision-making in clinical scenarios
McCradden et al [36]	NS	Potential for faster and more accurate analyses; ability to use more data	NS	There were concerns about privacy, commercial motives, and other risks and mixed views about explicit consent for research. Transparency is needed	It still requires human verification of computer-aided decisions	NS	Fear of losing human touch and skills from overreliance on machines

Study	Available on demand and user-friendly	Efficiency	Price	Lack of trust in data privacy	Lack of trust in patient safety	Lack of trust in technology	Concerns over full automation
McCradden et al [37]	NS	Predictive modeling performed on primary care health data and business analytics for primary care provider. AI has the potential to improve managerial and clinical decisions and processes, and this would be facilitated by common data standards	NS	Nonconsented use of health data is acceptable with disclosure and transparency. Selling health data should be prohibited. Some privacy health outcomes trade-off is acceptable	A few patients and caregivers felt that allocation of health resources should be done via computerized output, and a majority stated that it was inappropriate to delegate such decisions to a computer	NS	NS
Milne-Ives et al [23]	Easy to learn and use	Speed up the process of service delivery and performance. Respondents appreciated reminders and assistance in forming routines, chatbot agents in facilitating learning, and agents in providing accountability (eg, regular check-ins, follow-ups). Multimodal interactions (eg, voice, touch) were viewed positively	NS	NS	Unable to sufficiently encompass the real situational complexity. Electronic physician did not have the ability to go deep enough, provide access to other materials, or provide enough information	NS	NS
Nadarzynski et al [26]	Chatbots were perceived as a convenient tool that could facilitate the seeking of health information on the web	If free at the point of access, chatbots were seen as time-saving and useful platforms for triaging users to appropriate health care services	NS	Some participants were concerned about the ability of the chatbots to keep sensitive data secured and confidential. The level of anonymity offered by chatbots was viewed positively by several participants	Risk of harm from inaccurate or inadequate advice. Immature in performing a diagnosis but providing general health advice is acceptable	Uncertain about the quality, trustworthiness, and accuracy of the health information provided by chatbots	NS
Okolo et al [42]	NS	AI app would be able to perform some of the manual tasks and make the work of CHWs[c] more efficient, and help CHWs and patients in decision-making processes	NS	NS	Concerned over AI failures or misdiagnoses. The AI app might serve to reinforce the expertise of CHWs, improve patients' understanding of the diagnosis		AI would never completely replace health care workers because of the need for human interaction

Study	Available on demand and user-friendly	Efficiency	Price	Lack of trust in data privacy	Lack of trust in patient safety	Lack of trust in technology	Concerns over full automation
Palanica et al [27]	NS	Many physicians believed that chatbots would be most beneficial for administrative tasks such as scheduling physician appointments, locating health clinics, or providing medication information	NS	NS	Chatbots could be a risk to patients if they self-diagnose too often and did not accurately understand the diagnoses	NS	Chatbots alone are not able to provide effective care for all patients because of limited knowledge of personal factors
Prakash and Das [24]	Always available at the touch of a button and user-friendly	NS	The price of mental health chatbots could be a decisive factor in places with a poor health insurance system	Data privacy is a major barrier that prevents the adoption of mental health chatbots	Chatbots may be useful in managing mental health conditions but not good enough for complex problems. May even be more harmful to vulnerable patients with poor advice	Doubtful about reliability and functionality	NS
Scheetz et al [35]	NS	The top three potential advantages are improved patient access to disease screening; improved diagnostic confidence; and enhanced efficiency, that is, reduced time spent by specialists on monotonous tasks	NS	There were concerns over the divestment of health care to large technology and data companies	There were concerns over medical liability because of machine errors	AI would need to perform much more superior to the average specialist in screening and diagnosis	There is decreasing reliance on medical specialists for diagnosis and treatment advice
Stai et al [45]	NS	NS	Almost all (94%) participants were willing to pay for a review of medical imaging by an AI	NS	NS	Nearly equal trust in AI vs physician diagnoses; significantly more likely to trust an AI diagnosis of cancer over a physician's diagnosis	NS
Sun and Medaglia [40]	NS	NS	High treatment costs for patients but does not make profits for hospitals	Lack of trust toward AI-based decisions; unethical use of shared data	Doubts in the ability of AI to identify country-specific patient disease profiles	There were concerns over the lack of data integration; standards of data collection, format, and quality; algorithm opacity; and ability to read unstructured data	NS
Tam-Seto et al [38]	It could support those not currently accessing mental health services	It would address the perceived mental health service gap	NS	No assurance of users' privacy	Trust in the app, as it discloses that the app was informed by the Canadian military experience (credibility)	There were doubts over overall sustainability	NS

Study	Available on demand and user-friendly	Efficiency	Price	Lack of trust in data privacy	Lack of trust in patient safety	Lack of trust in technology	Concerns over full automation
Xiang et al [33]	NS	Health care workers prefer AI to alleviate daily repetitive work and improve outpatient guidance and consultation. The current auxiliary and partial substitution effects of AI are recognized by >90% of the public, and both groups have positive attitudes regarding AI development	NS	NS	Both health care and non–health care workers express more trust in real doctors than in AI	NS	A very small minority of health care and non–health care workers expect that full automation is likely to happen
Zhang et al [46]	NS	NS	NS	There were concerns about cybersecurity	NS	There were concerns about accuracy, reliability, quality, and trustworthiness of AI outputs, such as the predictions and recommended medical information	Supplementary service rather than a replacement of the professional health force is required for the AI to be particularly useful in helping patients to comprehend their physician's diagnosis

[a]AI: artificial intelligence.

[b]NS: not specified.

[c]CHW: community health care worker.

Availability and Ease of Use

Of the 26 studies, 3 (12%) studies highlighted the advantage of AI being constantly available without restrictions such as physical location, time, and access to a structured treatment [24,30,38]; 3 (12%) other studies also mentioned the appreciation of respondents for how an AI system could collect data remotely in a nonintrusive and user-friendly manner [23,24,28]. These studies mostly represented the perceptions of consumers and health care providers [24,30,38] (Multimedia Appendix 3). Only 4% (1/26) of studies did not mention the population characteristics [24].

Improves Efficiency and Reduces the Cost of Health Care Service Delivery

In all, 58% (15/26) of studies highlighted the potential of AI to improve the efficiency of health care service delivery in terms of remote monitoring [28], providing health-related reminders [23,28], increasing the speed and accuracy of health care processes (eg, consultation wait time, triaging, diagnosis, and managing medication refills) [26,29,30,35-37,44], facilitating care team communications, improving care accountability (eg, regular check-ins and follow-ups for information gathering) [23], and taking over repetitive manual tasks (eg, scheduling, patient education, and vital signs monitoring) [27]. Some respondents also appreciated the use of AI to provide a second opinion to physicians' diagnoses or evaluations [42,46]. Overall, 12% (3/26) of studies [24,34,45] discussed the potential cost-saving capacity of AI that influences AI acceptability, whereas 4% (1/26) mentioned that the provision of an AI service using IBM Watson caused patients to incur higher treatment costs that did not translate to profits for the hospital after factoring onboarding of the technology [40]. There was a good proportion of representation from the health care and IT staff (53.3%) [27-29,36,37,40,42,44] and those from the public, including patients (Multimedia Appendix 3). Only 4% (1/26) of the studies did not mention the population characteristics [24].

Concerns and Mitigation Strategies

Overview

Our findings highlight several concerns (Table 2) and mitigation strategies (Table 3).

Table 3. Needs and mitigation strategies of artificial intelligence (AI) in health care.

Study	Need for transparency, credibility, and regulation	Lack of personalization and customizability	Perceived empathy and personification	Design, user experience, and interconnectedness with other devices	Educating the public on AI capabilities
Abdi et al [28]	NS[a]	NS	NS	Implementing user-led design principles could facilitate the acceptability and uptake of these technologies	NS
Abdullah and Fakieh [29]	NS	NS	NS	NS	Most respondents had a general lack of AI knowledge (mean score 2.95 from 4) and were unaware of the advantages and challenges of AI applications in health care
Baldauf et al [30]	Need guarantee of anonymized transmission and analysis of personal health data of users	• Personalized explanation of analyses • Disease information • Treatment cost • Recommending physician's visit • Alternative Therapies • Prevention information • Treatment companion • Mental support • Objectivity and independence	Lack of personal face-to-face contact with a human expert	NS	NS
Castagno and Khalifa [43]	NS	NS	NS	NS	NS
Easton et al [31]	Needed clarity on whether the chatbot was a physician or an adviser	The system should allow personalization	The chatbot should be enriched by the ability to detect emotion (distress, fatigue, and irritation) in speech and nonverbal cues to build a therapeutic relationship between the agent and the patient	Personification of the chatbot should be emotionally expressive. Multi-modal interactions and interconnectedness with other consumer devices were suggested	NS
Gao et al [22]	NS	NS	NS	NS	NS
Griffin et al [25]	NS	NS	NS	Some older adults described limited use of smartphone, given the small screen or inability to keep track of it	NS
Kim [47]	NS	NS	NS	NS	NS
Laï et al [41]	Need for app regulation to create a more permissive regulatory framework; achieve confidentiality and respect for privacy	NS	NS	NS	NS

Study	Need for transparency, credibility, and regulation	Lack of personalization and customizability	Perceived empathy and personification	Design, user experience, and interconnectedness with other devices	Educating the public on AI capabilities
Li et al [32]	Credibility of the intelligent self-diagnosis system can be improved through transparency (eg, showing accuracy scores). State if doctors would accept information provided by AI	AI systems may provide more specific, personalized information and advice	NS	NS	NS
Liu et al [34]	NS	NS	NS	NS	NS
Liu et al [39]	NS	NS	NS	NS	NS
Liyanage et al [44]	NS	NS	NS	NS	NS
McCradden et al [36]	Need for transparency on how and by whom their data were used	NS	NS	NS	NS
McCradden et al [37]	Need for transparency, disclosure, reparations, deidentification of data, and use within trusted institutions	NS	NS	NS	NS
Milne-Ives et al [23]	NS	Need more customization or availability of feature options (eg, preformatted or free-text options)	Need for greater interactivity or relational skills in conversational agents. Respondents liked that the agent had a personality and showed empathy, which improves personal connection. Others had difficulty in empathizing with the agent or reported disliking its limited conversation and responses	Interaction was too long, the use of nonverbal expressions by the avatar was not appealing, and there was a lack of clarity regarding the aim of the chatbot. Better integration of the agent with electronic health record systems (for a virtual physician) or health care providers (for an asthma self-management chatbot) would be useful	NS
Nadarzynski et al [26]	Need to increase transparency of information source	NS	Lack of empathy and inability of chatbots to understand more emotional issues, especially in mental health. The responses given by chatbots were seen as depersonalized, cold, and inhuman. They were perceived as inferior to physician consultation, although anonymity could facilitate the disclosure of more intimate or uncomfortable aspects to do with health	NS	There was a general lack of familiarity and understanding of health chatbots among participants
Okolo et al [42]	NS	NS	NS	NS	NS
Palanica et al [27]	NS	NS	Many physicians believed that chatbots cannot display human emotion	NS	NS

Study	Need for transparency, credibility, and regulation	Lack of personalization and customizability	Perceived empathy and personification	Design, user experience, and interconnectedness with other devices	Educating the public on AI capabilities
Prakash and Das [24]	NS	There were user input restrictions during chatbot conversations where the chatbot forced the users to respond to a list of choices	• Mixed findings on perceived empathy. Some users perceived the chatbot to be warm and friendly, whereas others found it to be unsympathetic and rude • Mixed findings on preference for a lifelike chatbot—some felt it a little creepy and weird • The nonjudgmental nature of chatbots is a strong motivator of adoption. It should respond spontaneously in a contingent, humanlike manner	NS	NS
Scheetz et al [35]	NS	NS	NS	NS	A minority (13.8%) of the participants felt that the specialist training colleges were adequately prepared for the introduction of AI into clinical practice. Education was identified as a priority to prepare clinicians for the implementation of AI in health care
Stai et al [45]	NS	NS	NS	NS	NS
Sun and Medaglia [40]	NS	NS	NS	NS	Insufficient knowledge on values and advantages of AI technology; unrealistic expectations toward AI technology
Tam-Seto et al [38]	NS	NS	NS	NS	Managing the public's expectations of the capabilities of such an app
Xiang et al [33]	NS	NS	NS	NS	More than 90% of health care workers expressed a willingness to devote time to learning about AI and participating in AI research

Study	Need for transparency, credibility, and regulation	Lack of personalization and customizability	Perceived empathy and personification	Design, user experience, and interconnectedness with other devices	Educating the public on AI capabilities
Zhang et al [46]	Majority of participants expressed the need to increase system transparency by explaining how the AI arrived at its conclusion	• Need more personalized and actionable information • AI should be enhanced with features that can help to recommend personalized questions to ask physicians	Concerns over lack of empathy	NS	NS

[a]NS: not specified.

Lack of Trust

Data Privacy

In all, 58% (15/26) of studies described the respondents' lack of trust regarding how their personal data will be collected (eg, unknowingly through voice-activated devices) and handled (eg, by whom and how) [22,24-26,28,30,31,35,36,38,40,41,43,46]. However, 4% (1/26) of the studies reported no concerns regarding data sharing. This could be because of the respondents being chronic obstructive pulmonary disease patients who may have been used to their data being shared for clinical decision-making purposes [31]. Potential mitigation strategies suggested were to guarantee anonymity [26] and increase transparency in how the collected data will be used (eg, by which third party and how) [24,37]. There was a good proportion of representation from the general public, including patients (53.3%) [22,24-26,30,31,37,38,46] and health care providers and IT staff (Multimedia Appendix 3).

Patient Safety

Of the 26 studies, 21 (81%) discussed the respondents' lack of trust in an AI to ensure patient safety while performing its tasks, especially regarding providing accurate information on rare conditions or unexpected situations [22-27,29-42,44]. Other concerns were regarding the credibility of AI-based recommendations (eg, whether it was validated by medical professionals) [30,32], maturity in the technology to provide safe and realistic recommendations [22,25], medical liability from the risk of medical errors and bias [26,35,36,44], secondary effects of AI-based diagnoses such as insurance claims [44], and miscommunications [26]. The potential mitigation strategies suggested were the provision of AI-specific regulations [30,31,41], transparency in its credibility, how a recommendation is derived (eg, showing who developed the system and the system reasoning and reliability based on information source and personal information), and its accuracy [32,38]. In contrast, 4% (1/26) of studies reported that the respondents were confident that the AI would outperform human clinical diagnoses because of higher accuracy and lower human errors [39]. Most respondents accepted AI in providing general health advice to minor ailments. Most of the responses represented the voices of the public, including patients (66.6%) [22-26,30-32,34,35,37-40] (Multimedia Appendix 3).

Technology

Of the 26 studies, 6 (23%) studies discussed the participants' lack of trust in the maturity of AI technology in providing reliable and accurate information to support health-related predictions and recommendations [24,26,35,38,40,46]. This could be related to concerns over the lack of integration and synthesis of information from various sources, standardization of data collection, and the overall sustainability of AI-assisted health care service delivery [40,45]. However, 8% (2/26) of studies reported that respondents had similar trust in AI as compared with a human physician's diagnoses [28,45]. Possible mitigation strategies include increasing system transparency and reporting system accuracies [26,46]. Only 8% (2/26) of studies represented the voices of health care and IT staff [35,40,49] (Multimedia Appendix 3).

Potential Impacts of Full Automation

In all, 46% (12/26) of studies discussed the perceptions of respondents on the possibility and impacts of full automation on the health care industry, especially in terms of diagnoses, all of which reported that it is unlikely that AI will completely replace health care professionals [22, 27, 29, 30, 33, 35, 36, 39, 42-44, 46]. This could largely be because of the immaturity of AI technology and its limitations in providing human-like interactions (which build trust) [27]. Instead, many patients preferred a combination of both AI and human physicians in diagnoses to achieve a more accurate and comprehensive evaluation [30,39]. Most of the responses represented the voices of health care and IT staff (58.3%) [27,29,35,36,42-44] (Multimedia Appendix 3).

Needs to Improve Adoption of AI in Health Care

Besides the needs highlighted to mitigate the concerns, several additional features were found to potentially improve the adoption of AI in health care (Table 3).

Enhance Personalization and Customizability

Of the 26 studies, 6 (23%) studies discussed the need for AI to personalize information such as the explanation of diagnoses, recommendations, patient education, and even pertinent questions or issues to raise to their physicians [23,24,30-32,46]. Some studies also mentioned the need to customize chatbot features according to user preferences (for fixed options or free-texts) [23,24].

Enhance Empathy and Personification of AI-Enabled Chatbots and Avatars

In all, 27% (7/26) of studies highlighted the respondents' concern over the lack of empathy, which is a crucial element of human interaction to build trust between service providers and consumers. However, empathy must be displayed tactfully in verbal and nonverbal expressions such that it does not appear to be "creepy and weird," especially in populations with mental health issues [24]. Personification was also emphasized to increase the relatability, connection, and appeal to interact with the chatbot or avatar [23]. Perceived anonymity in interacting with the chatbot was also highlighted to assist in communication regarding sensitive topics [26].

Enhance User Experience, Design, and Interconnectedness With Other Devices

Overall, 15% (4/26) of studies described the need to improve user experience to increase user engagement with AI [23,25,28,31]. Strategies include needs-based interaction timing, the use of suitable verbal and nonverbal expressions, interconnectedness with other information sources (eg, electronic health record), apps (eg, calendar), and devices (eg, smart home technology–enabled devices).

Educate the Public on AI Capabilities

Of the 26 studies, 6 (23%) studies highlighted the lack of public and clinical awareness on the capabilities of AI in health care, of which the majority of the respondents expressed their willingness to learn [26,29,33,35,38,40]. A better understanding of the advantages and disadvantages of AI in health care could enhance the health care service delivery efficiency while balancing the expectations from it.

Discussion

Principal Findings

On the basis of the 26 articles included in this scoping review, we identified the perceptions and needs of various populations in the use of AI for general, primary, and community health care; chronic diseases self-management; self-diagnosis; mental health; and diagnostic procedures. However, the use of AI in health care remains challenged by the common perceptions, concerns, and unmet needs of various stakeholders such as patients, health care professionals, governmental or legal regulatory bodies, software developers, and industrial providers. Simply introducing AI into health care systems without understanding the needs of stakeholders will not lead to a sustainable change [50].

Our results showed that, similar to most ITs, AI was generally favored for its on-demand availability, ease of use, potential to improve efficiency, and reduce the cost of health care service delivery. These features could enhance patients' compliance to health care treatments and recommendations that may be inaccessible or inconvenient. For example, patients are traditionally required to commit to a physician's consultative appointment that could be relatively inflexible because of a long list of patients, and one could be forced to skip the consultation because of a conflict in their schedule. AI confers the benefit of information collection and dissemination beyond the constraints of time and place, which have been shown to improve medication adherence through an AI-based smartphone app [51] and diet and exercise adherence through an AI-based virtual health assistant [52]. Our findings also demonstrated that AI is valued for its potential to speed up health care processes such as diagnosis, waiting time, communication with care teams, decisional support, and other routine tasks (eg, progress monitoring) that can be automated. This increase in service delivery efficiency frees up time and resources for clinicians to focus on tasks that involve more unexpected variabilities such as dealing with rare disease management and interacting with patients, thereby reducing the risk of burnout, job dissatisfaction, and manpower shortage [53].

Although our findings showed high rates of acceptability, concerns were raised about the lack of trust (in data privacy, patient safety, and technology maturity) and the impacts of AI-driven automation on health care job security and health care services. Ethical controversies surrounding the use of AI in health care have been long-standing. Although there are increasingly more regulatory guidelines available, such as those developed by the World Health Organization [54] and the European Union [55], the use of AI in health care remains debatable because of the challenges in ensuring data privacy and proper data use [56]. This is especially true when data collection modes are conducted through third-party apps, such as Facebook Messenger (Meta Platforms), of which privacy policies are governed by technology companies and not health care institutions [24]. Moreover, although there are privacy and security precautionary measures, the increasing reports of data leaks and vulnerabilities in electronic medical record databases erode population trust. Future security and transparency measures could consider the use of blockchain technology, and privacy laws should be properly delineated and transparent [57].

This review also found the need to enhance the personalization and customizability of information provided by AI, the incorporation of empathy and personification in AI-based conversational agents, the user experience through better design and interconnectedness with other devices and systems, and the need to educate the public on AI capabilities. Concerning personalized health care, reports generated by AI should be integrated and explained in accordance with each individual's demographic and clinical profile to facilitate self-management [46]. We also identified the need for AI to not only assist in the understanding of patients' medical condition but also the provision of relevant treatment options and personalized recommendations with intuitive actions provided (eg, a button to call an ambulance when deemed necessary by the AI) [31]. This coincides with existing studies that highlight the predictive power of AI in providing support to preventive disease onset or deterioration through interventions tailored according to user preferences [58]. For example, AI has been used to provide just-in-time adaptive interventions that prompt users to perform healthy behavior changes (eg, healthy diet and exercise and smoking cessation) based on constant data collection of their behaviors and preferences [49]. However, the data collection of users' behavioral or clinical information should also consider the customizability of input options (eg, providing predefined

options or allowing for free-text input) to enhance the usability and adoption of such systems, depending on user preferences [24]. Personification of AI-based conversational agents to express human-like identity, personality, empathy, and emotions was also highlighted as an area of improvement to enhance human-chatbot interactions and eventually user adoption [59]. It was also important for the AI systems to be accessible through various devices (eg, tablets, televisions, laptops, and smart home appliances) and modes (eg, text and speech) for the convenience of information consumption and data collection. Finally, our findings suggest a need to address the knowledge deficit in the definition, capacity, and functions of AI. This could be done by cultivating AI literacy and exposure from childhood [60] and incorporating the AI curriculum in health care training and upgrading courses [61].

Overall, our study findings are consistent with well-established theories such as the Technology Acceptance Model, of which the second version proposed by Venkatesh and Davis [62] posits that technology acceptance is strongly associated with the perceived usefulness and perceived ease of use, which are influenced by subjective norms, images, job relevance, output quality, result demonstrability, experience, and voluntariness [63]. Therefore, to enhance the acceptability of AI in health care applications, its perceived usefulness over and above the current standard practices such as capacity to increase service delivery efficiency and community-based self-diagnostic accuracy should be emphasized. Such messages should be designed to be relevant to the individual and organizational adopters of a social system through various communication channels and change agents (ie, gatekeepers and opinion leaders). Such messages should be persuasive to spark five stages of adoption, namely, knowledge, persuasion, decision, implementation, and confirmation, known as the diffusion of innovation theory by Rogers [64]. Different strategies are also needed to correspond with the different categories of adopters, namely, the innovators, early adopters, early majority, late majority, and laggards. Different rates of technology adoption are associated with one's risk tolerance related to higher social economic status, education level, and financial stability [65]. An example is the case of AI adoption in chronic disease early detection and management in the United Arab Emirates. Success was attributed to the *managerial, organizational, operational, and IT infrastructure factors* that contribute to the factors of the Technology Acceptance Model [66]. However, advanced technologies such as AI continue to be relatively expensive and require eHealth literacy, which may widen the digital divide, and therefore the data divide and health disparity among societies. According to a report published in *The Lancet*, the internet remains inaccessible to approximately 50% of the global population because of a digital divide [67]. In addition, there are specific guidelines on the implementation of AI in health care service delivery, such as the quality of data and certification of AI systems, which may deter adoption [68].

Limitations

This study had several limitations. First, only articles written in English were retrieved, possibly limiting the comprehensiveness of our findings. However, we conducted a search on Google Scholar to supplement the electronic database search for more relevant papers. Second, the studies were largely heterogeneous in their study designs, research aims, and data collection methods. Third, there were limited studies on the perceptions of AI and clinical researchers who could provide outlooks on the perceptions of the general public. Finally, the public's perceptions of AI in health care may be limited by their knowledge of the definitions and capabilities of AI, as highlighted in our findings that there is a need to enhance the public's knowledge on AI. Therefore, the priority or importance of each perception and need could not be evaluated. The inclusion of articles based on our definition of AI could also have limited the scope of this study. Studies that considered different definitions of AI may have been excluded.

Recommendations for Future Design and Research

This study highlighted the perceptions and needs of AI to enhance its adoption in health care. However, one major challenge lies in the extent to which AI is tailored according to each individual's unique preference, and if such preferences are largely varied, how data can be aggregated for analyses and applicability in specific health care applications. Therefore, future studies that use AI should not only consider the issues raised in this study but also clarify the applicability in their applications and target population. A prior needs-based analysis is recommended before the development of AI systems.

Conclusions

Although AI is valued for its 24/7 availability in health care service delivery, ease of use, and capacity to improve health care service provision efficiency, concerns over trust in data privacy, information credibility, and technological maturity remain. Although several mitigation strategies such as enhancing transparency over predictive accuracy and information sources were identified, other areas of improvement were also highlighted. Future studies and AI development should consider the points raised in this study to enhance the adoption and enhancement of AI to improve health care service delivery.

Acknowledgments

This research was supported by the National University Health System Internal Grant Funding under grant NUHSRO/2021/063/RO5+6/FMPCHSRG-Mar21/01 and the National Research Foundation, Singapore, under its Strategic Capabilities Research Centres Funding Initiative. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author or authors and do not reflect the views of the National University Health System or the National Research Foundation, Singapore.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) checklist.
[DOCX File , 16 KB - jmir_v24i1e32939_app1.docx]

Multimedia Appendix 2
Database search details.
[DOCX File , 14 KB - jmir_v24i1e32939_app2.docx]

Multimedia Appendix 3
Study characteristics.
[DOCX File , 21 KB - jmir_v24i1e32939_app3.docx]

Multimedia Appendix 4
Acceptability of artificial intelligence use in health care.
[DOCX File , 16 KB - jmir_v24i1e32939_app4.docx]

References

1. Panch T, Szolovits P, Atun R. Artificial intelligence, machine learning and health systems. J Glob Health 2018 Dec;8(2):020303 [FREE Full text] [doi: 10.7189/jogh.08.020303] [Medline: 30405904]
2. Chew HS, Ang WH, Lau Y. The potential of artificial intelligence in enhancing adult weight loss: a scoping review. Public Health Nutr 2021 Jun;24(8):1993-2020 [FREE Full text] [doi: 10.1017/S1368980021000598] [Medline: 33592164]
3. Panch T, Pearson-Stuttard J, Greaves F, Atun R. Artificial intelligence: opportunities and risks for public health. Lancet Digit Health 2019 May;1(1):13-14. [doi: 10.1016/s2589-7500(19)30002-0]
4. Arora A. Conceptualising artificial intelligence as a digital healthcare innovation: an introductory review. Med Devices (Auckl) 2020 Aug 20;13:223-230 [FREE Full text] [doi: 10.2147/MDER.S262590] [Medline: 32904333]
5. Dilsizian SE, Siegel EL. Artificial intelligence in medicine and cardiac imaging: harnessing big data and advanced computing to provide personalized medical diagnosis and treatment. Curr Cardiol Rep 2014 Jan;16(1):441. [doi: 10.1007/s11886-013-0441-8] [Medline: 24338557]
6. Hashimoto D, Rosman G, Rus D, Meireles O. Artificial intelligence in surgery: promises and perils. Ann Surg 2018 Jul;268(1):70-76 [FREE Full text] [doi: 10.1097/SLA.0000000000002693] [Medline: 29389679]
7. Stein N, Brooks K. A fully automated conversational artificial intelligence for weight loss: longitudinal observational study among overweight and obese adults. JMIR Diabetes 2017 Nov 01;2(2):e28 [FREE Full text] [doi: 10.2196/diabetes.8590] [Medline: 30291087]
8. Sahoo D, Hao W, Ke S, Xiongwei W, Le H, Achananuparp P, et al. FoodAI: food image recognition via deep learning for smart food logging. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2019 Presented at: 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; Aug 4 - 8, 2019; Anchorage AK USA. [doi: 10.1145/3292500.3330734]
9. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
10. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
11. Beam AL, Kohane IS. Translating artificial intelligence into clinical care. J Am Med Assoc 2016 Dec 13;316(22):2368-2369. [doi: 10.1001/jama.2016.17217] [Medline: 27898974]
12. Lu Y, Xiao Y, Sears A, Jacko JA. A review and a framework of handheld computer adoption in healthcare. Int J Med Inform 2005 Jun;74(5):409-422. [doi: 10.1016/j.ijmedinf.2005.03.001] [Medline: 15893264]
13. Or CK, Karsh B. A systematic review of patient acceptance of consumer health information technology. J Am Med Informatics Assoc 2009 Jul 01;16(4):550-560. [doi: 10.1197/jamia.m2888]
14. Yusif S, Soar J, Hafeez-Baig A. Older people, assistive technologies, and the barriers to adoption: a systematic review. Int J Med Inform 2016 Oct;94:112-116. [doi: 10.1016/j.ijmedinf.2016.07.004] [Medline: 27573318]
15. Ethics guidelines for trustworthy AI. European Commission. URL: https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html [accessed 2021-12-28]
16. Ting DS, Liu Y, Burlina P, Xu X, Bressler NM, Wong TY. AI for medical imaging goes deep. Nat Med 2018 May;24(5):539-540. [doi: 10.1038/s41591-018-0029-3] [Medline: 29736024]

17. Quiroz JC, Laranjo L, Kocaballi AB, Berkovsky S, Rezazadegan D, Coiera E. Challenges of developing a digital scribe to reduce clinical documentation burden. NPJ Digit Med 2019;2:114 [FREE Full text] [doi: 10.1038/s41746-019-0190-1] [Medline: 31799422]
18. Lysaght T, Lim HY, Xafis V, Ngiam KY. Ai-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research. Asian Bioeth Rev 2019 Sep;11(3):299-314 [FREE Full text] [doi: 10.1007/s41649-019-00096-0] [Medline: 33717318]
19. Delipetrev B, Tsinaraki C, Kostic U. Historical Evolution of Artificial Intelligence. Luxembourg: Publications Office of the European Union; 2020.
20. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
21. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
22. Gao S, He L, Chen Y, Li D, Lai K. Public perception of artificial intelligence in medical care: content analysis of social media. J Med Internet Res 2020 Jul 13;22(7):e16649 [FREE Full text] [doi: 10.2196/16649] [Medline: 32673231]
23. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
24. Prakash A, Das S. Intelligent conversational agents in mental healthcare services: a thematic analysis of user perceptions. Pacific Asia J Assoc Inf Syst 2020:1-34 [FREE Full text]
25. Griffin A, Xing Z, Mikles S, Bailey S, Khairat S, Arguello J, et al. Information needs and perceptions of chatbots for hypertension medication self-management: a mixed methods study. JAMIA Open 2021 Apr;4(2):ooab021 [FREE Full text] [doi: 10.1093/jamiaopen/ooab021] [Medline: 33898936]
26. Nadarzynski T, Miles O, Cowie A, Ridge D. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: a mixed-methods study. Digit Health 2019;5:2055207619871808 [FREE Full text] [doi: 10.1177/2055207619871808] [Medline: 31467682]
27. Palanica A, Flaschner P, Thommandram A, Li M, Fossat Y. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. J Med Internet Res 2019 Apr 05;21(4):e12887 [FREE Full text] [doi: 10.2196/12887] [Medline: 30950796]
28. Abdi S, Witte LD, Hawley M. Exploring the potential of emerging technologies to meet the care and support needs of older people: a delphi survey. Geriatrics (Basel) 2021 Feb 13;6(1):19 [FREE Full text] [doi: 10.3390/geriatrics6010019] [Medline: 33668557]
29. Abdullah R, Fakieh B. Health care employees' perceptions of the use of artificial intelligence applications: survey study. J Med Internet Res 2020 May 14;22(5):e17620 [FREE Full text] [doi: 10.2196/17620] [Medline: 32406857]
30. Baldauf M, Fröehlich P, Endl R. Trust me, I'm a doctor – user perceptions of AI-driven apps for mobile health diagnosis. In: Proceedings of the 19th International Conference on Mobile and Ubiquitous Multimedia. 2020 Presented at: MUM 2020: 19th International Conference on Mobile and Ubiquitous Multimedia; Nov 22 - 25, 2020; Essen Germany. [doi: 10.1145/3428361.3428362]
31. Easton K, Potter S, Bec R, Bennion M, Christensen H, Grindell C, et al. A virtual agent to support individuals living with physical and mental comorbidities: co-design and acceptability testing. J Med Internet Res 2019 May 30;21(5):e12996 [FREE Full text] [doi: 10.2196/12996] [Medline: 31148545]
32. Li W, Fan X, Zhu H, Wu J, Teng D. Research on the influencing factors of user trust based on artificial intelligence self diagnosis system. In: Proceedings of the ACM Turing Celebration Conference. 2020 Presented at: ACM Turing Celebration Conference; May 22 - 24, 2020; Hefei China. [doi: 10.1145/3393527.3393561]
33. Xiang Y, Zhao L, Liu Z, Wu X, Chen J, Long E, et al. Implementation of artificial intelligence in medicine: status analysis and development suggestions. Artif Intell Med 2020 Jan;102:101780. [doi: 10.1016/j.artmed.2019.101780] [Medline: 31980086]
34. Liu T, Tsang W, Xie Y, Tian K, Huang F, Chen Y, et al. Preferences for artificial intelligence clinicians before and during the COVID-19 pandemic: discrete choice experiment and propensity score matching study. J Med Internet Res 2021 Mar 02;23(3):e26997 [FREE Full text] [doi: 10.2196/26997] [Medline: 33556034]
35. Scheetz J, Rothschild P, McGuinness M, Hadoux X, Soyer HP, Janda M, et al. A survey of clinicians on the use of artificial intelligence in ophthalmology, dermatology, radiology and radiation oncology. Sci Rep 2021 Mar 04;11(1):5193 [FREE Full text] [doi: 10.1038/s41598-021-84698-5] [Medline: 33664367]
36. McCradden MD, Baba A, Saha A, Ahmad S, Boparai K, Fadaiefard P, et al. Ethical concerns around use of artificial intelligence in health care research from the perspective of patients with meningioma, caregivers and health care providers: a qualitative study. CMAJ Open 2020;8(1):90-95 [FREE Full text] [doi: 10.9778/cmajo.20190151] [Medline: 32071143]
37. McCradden MD, Sarker T, Paprica PA. Conditionally positive: a qualitative study of public perceptions about using health data for artificial intelligence research. BMJ Open 2020 Oct 28;10(10):e039798 [FREE Full text] [doi: 10.1136/bmjopen-2020-039798] [Medline: 33115901]

38. Tam-Seto L, Wood VM, Linden B, Stuart H. Perceptions of an AI-supported mobile app for military health in the Canadian armed forces. Milit Behav Health 2020 Nov 13;9(3):247-254. [doi: 10.1080/21635781.2020.1838364]
39. Liu T, Tsang W, Huang F, Lau OY, Chen Y, Sheng J, et al. Patients' preferences for artificial intelligence applications versus clinicians in disease diagnosis during the SARS-CoV-2 pandemic in China: discrete choice experiment. J Med Internet Res 2021 Feb 23;23(2):e22841 [FREE Full text] [doi: 10.2196/22841] [Medline: 33493130]
40. Sun TQ, Medaglia R. Mapping the challenges of Artificial Intelligence in the public sector: evidence from public healthcare. Govern Inform Q 2019 Apr;36(2):368-383. [doi: 10.1016/j.giq.2018.09.008]
41. Laï M, Brian M, Mamzer M. Perceptions of artificial intelligence in healthcare: findings from a qualitative survey study among actors in France. J Transl Med 2020 Jan 09;18(1):14 [FREE Full text] [doi: 10.1186/s12967-019-02204-y] [Medline: 31918710]
42. Okolo C, Kamath S, Dell N, Vashistha A. “It cannot do all of my work”: community health worker perceptions of AI-enabled mobile health applications in rural India. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan. [doi: 10.1145/3411764.3445420]
43. Castagno S, Khalifa M. Perceptions of artificial intelligence among healthcare staff: a qualitative survey study. Front Artif Intell 2020 Oct 21;3:578983 [FREE Full text] [doi: 10.3389/frai.2020.578983] [Medline: 33733219]
44. Liyanage H, Liaw S, Jonnagaddala J, Schreiber R, Kuziemsky C, Terry AL, et al. Artificial intelligence in primary health care: perceptions, issues, and challenges. Yearb Med Inform 2019 Aug;28(1):41-46 [FREE Full text] [doi: 10.1055/s-0039-1677901] [Medline: 31022751]
45. Stai B, Heller N, McSweeney S, Rickman J, Blake P, Vasdev R, et al. Public perceptions of artificial intelligence and robotics in medicine. J Endourol 2020 Oct;34(10):1041-1048 [FREE Full text] [doi: 10.1089/end.2020.0137] [Medline: 32611217]
46. Zhang Z, Citardi D, Wang D, Genc Y, Shan J, Fan X. Patients' perceptions of using artificial intelligence (AI)-based technology to comprehend radiology imaging data. Health Informatics J 2021;27(2):14604582211011215 [FREE Full text] [doi: 10.1177/14604582211011215] [Medline: 33913359]
47. Kim H. An analysis of the need for aid tools in dementia patients: focusing on the normal elderly, dementia patients, and caregivers of dementia patients. Ind J Public Health Res Develop 2019;10(11):4399. [doi: 10.5958/0976-5506.2019.04300.6]
48. Kim S, Kim J, Badu-Baiden F, Giroux M, Choi Y. Preference for robot service or human service in hotels? Impacts of the COVID-19 pandemic. Int J Hospitality Manag 2021 Feb;93:102795. [doi: 10.1016/j.ijhm.2020.102795]
49. Nahum-Shani I, Smith S, Spring B, Collins L, Witkiewitz K, Tewari A, et al. Just-in-Time Adaptive Interventions (JITAIs) in mobile health: key components and design principles for ongoing health behavior support. Ann Behav Med 2018 May 18;52(6):446-462 [FREE Full text] [doi: 10.1007/s12160-016-9830-8] [Medline: 27663578]
50. Panch T, Mattie H, Celi LA. The "inconvenient truth" about AI in healthcare. NPJ Digit Med 2019;2:77 [FREE Full text] [doi: 10.1038/s41746-019-0155-4] [Medline: 31453372]
51. Roosan D, Chok J, Karim M, Law AV, Baskys A, Hwang A, et al. Artificial intelligence-powered smartphone app to facilitate medication adherence: protocol for a human factors design study. JMIR Res Protoc 2020 Nov 09;9(11):e21659 [FREE Full text] [doi: 10.2196/21659] [Medline: 33164898]
52. Davis CR, Murphy KJ, Curtis RG, Maher CA. A process evaluation examining the performance, adherence, and acceptability of a physical activity and diet artificial intelligence virtual health assistant. Int J Environ Res Public Health 2020 Dec 07;17(23):9137 [FREE Full text] [doi: 10.3390/ijerph17239137] [Medline: 33297456]
53. Meskó B, Hetényi G, Győrffy Z. Will artificial intelligence solve the human resource crisis in healthcare? BMC Health Serv Res 2018 Jul 13;18(1):545 [FREE Full text] [doi: 10.1186/s12913-018-3359-4] [Medline: 30001717]
54. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance. Geneva: World Health Organization; 2021.
55. White paper on Artificial Intelligence: a European approach to excellence and trust. European Commission. 2020. URL: https://ec.europa.eu/info/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en [accessed 2021-12-28]
56. Lee D, Yoon SN. Application of artificial intelligence-based technologies in the healthcare industry: opportunities and challenges. Int J Environ Res Public Health 2021 Jan 01;18(1):271 [FREE Full text] [doi: 10.3390/ijerph18010271] [Medline: 33401373]
57. Wahl B, Cossy-Gantner A, Germann S, Schwalbe NR. Artificial intelligence (AI) and global health: how can AI contribute to health in resource-poor settings? BMJ Glob Health 2018;3(4):e000798 [FREE Full text] [doi: 10.1136/bmjgh-2018-000798] [Medline: 30233828]
58. Shaban-Nejad A, Michalowski M, Buckeridge D. Health intelligence: how artificial intelligence transforms population and personalized health. NPJ Digit Med 2018 Oct 2;1:53 [FREE Full text] [doi: 10.1038/s41746-018-0058-9] [Medline: 31304332]
59. Chaves AP, Gerosa MA. How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. Int J Hum Comput Interact 2020 Nov 08;37(8):729-758. [doi: 10.1080/10447318.2020.1841438]
60. Teaching tech to talk: K-12 conversational artificial intelligence literacy curriculum and development tools. arXiv. 2020. URL: https://arxiv.org/abs/2009.05653 [accessed 2021-12-28]

61. Wood EA, Ange BL, Miller DD. Are we ready to integrate artificial intelligence literacy into medical school curriculum: students and faculty survey. J Med Educ Curric Dev 2021 Jun 23;8:23821205211024078 [FREE Full text] [doi: 10.1177/23821205211024078] [Medline: 34250242]
62. Venkatesh V, Davis FD. A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag Sci 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
63. Marangunić N, Granić A. Technology acceptance model: a literature review from 1986 to 2013. Univ Access Inf Soc 2014 Feb 16;14(1):81-95. [doi: 10.1007/s10209-014-0348-1]
64. Kaminski J. Diffusion of innovation theory. Can J Nursing Inf 2011;6(2):1-6. [doi: 10.1097/cin.0000000000000072]
65. Kwon H, Chidambaram L. A test of the technology acceptance model: the case of cellular telephone adoption. In: Proceedings of the 33rd Annual Hawaii International Conference on System Sciences. 2000 Presented at: 33rd Annual Hawaii International Conference on System Sciences; Jan 7, 2000; Maui, HI, USA. [doi: 10.1109/hicss.2000.926607]
66. Alhashmi S, Salloum S, Mhamdi C. Implementing artificial intelligence in the United Arab Emirates healthcare sector: an extended technology acceptance model. Int J Inf Technol Lang Stud 2019:27-42 [FREE Full text]
67. Makri A. Bridging the digital divide in health care. Lancet Digit Health 2019 Sep;1(5):204-205. [doi: 10.1016/s2589-7500(19)30111-6]
68. Thinking on its own: AI in the NHS. Reform Research Trust. URL: https://reform.uk/research/thinking-its-own-ai-nhs [accessed 2021-12-28]

Abbreviations

AGI: artificial general intelligence
AI: artificial intelligence
ANI: artificial narrow intelligence
IT: information technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by A Mavragani; submitted 16.08.21; peer-reviewed by N Tom, K Ludlow, S Hong; comments to author 04.10.21; revised version received 08.11.21; accepted 03.12.21; published 14.01.22.

Please cite as:
Chew HSJ, Achananuparp P
Perceptions and Needs of Artificial Intelligence in Health Care to Increase Adoption: Scoping Review
J Med Internet Res 2022;24(1):e32939
URL: https://www.jmir.org/2022/1/e32939
doi:10.2196/32939
PMID:35029538

Review

Risk of Accidents or Chronic Disorders From Improper Use of Mobile Phones: A Systematic Review and Meta-analysis

Xinxi Cao[1,2*], PhD; Yangyang Cheng[1*], MSc; Chenjie Xu[1], PhD; Yabing Hou[1], MSc; Hongxi Yang[1], PhD; Shu Li[1], PhD; Ying Gao[3], PhD; Peng Jia[4,5*], PhD; Yaogang Wang[1*], PhD

[1]School of Public Health, Tianjin Medical University, Tianjin, China

[2]School of Public Administration, Nanjing Normal University, Nanjing, China

[3]Tianjin Medical University General Hospital, Tianjin Medical University, Tianjin, China

[4]School of Resource and Environmental Sciences, Wuhan University, Wuhan, China

[5]International Institute of Spatial Lifecourse Epidemiology (ISLE), Wuhan University, Wuhan, China

[*]these authors contributed equally

Corresponding Author:
Yaogang Wang, PhD
School of Public Health
Tianjin Medical University
No 22 Qixiangtai Road
Heping District
Tianjin, 300070
China
Phone: 86 13820046130
Email: wyg@tmu.edu.cn

Abstract

Background: Mobile phone use has brought convenience, but the long or improper use of mobile phones can cause harm to the human body.

Objective: We aimed to assess the impact of improper mobile phone use on the risks of accidents and chronic disorders.

Methods: We systematically searched in PubMed, EMBASE, Cochrane, and Web of Science databases for studies published prior to April 5, 2019; relevant reviews were also searched to identify additional studies. A random-effects model was used to calculate the overall pooled estimates.

Results: Mobile phone users had a higher risk of accidents (relative risk [RR] 1.37, 95% CI 1.22 to 1.55). Long-term use of mobile phones increased accident risk relative to nonuse or short-term use (RR 2.10, 95% CI 1.63 to 2.70). Compared with nonuse, mobile phone use resulted in a higher risk for neoplasms (RR 1.07, 95% CI 1.01 to 1.14), eye diseases (RR 2.03, 95% CI 1.27 to 3.23), mental health disorders (RR 1.16, 95% CI 1.02 to 1.32), and headaches (RR 1.25, 95% CI 1.18 to 1.32); the pooled risk of other chronic disorders was 1.20 (95% CI 0.90 to 1.59). Subgroup analyses also confirmed the increased risk of accidents and chronic disorders.

Conclusions: Improper use of mobile phones can harm the human body. While enjoying the convenience brought by mobile phones, people have to use mobile phones properly and reasonably.

(J Med Internet Res 2022;24(1):e21313) doi:10.2196/21313

KEYWORDS

cell phone; mobile phone; accident; neoplasm; radiation

Introduction

In the first quarter of 2019, the number of mobile phone users reached 7.9 billion, with an increase of approximately 2% year-on-year [1]. China had the most net additions during this quarter (30 million), followed by Nigeria (5 million), and the Philippines (4 million). In addition, it was predicted that the worldwide mobile phone market would reach 1.5 billion shipment units by the end of 2019 and that the pending arrival of 5G would attract more phone users by 2020 or 2021 [2]. Although mobile phones facilitate people's daily lives and provide effective auxiliary means for the treatment and

management of diseases [3-5], the health hazards potentially caused by using mobile phones are also a growing concern.

Although many countries and regions have passed laws prohibiting the use of mobile phones while driving, the number of reported traffic accidents caused by using mobile phones while driving has been increasing in recent years [6,7]. Nearly one-quarter of all traffic accidents in the United Kingdom in 2013 were caused by drivers using phones while driving. In addition, harm may be caused to the ears, parotid glands, and indirect brain areas during mobile phone usage [8-10]. Some in vivo or in vitro [11], simulator [12], or real-world studies [13] have been carried out to test the effects between human body and mobile phone use. There is currently no consensus on the use of mobile phones and chronic disorders, especially with respect to neoplasms, because results conflict. Mobile phone radiation has been classified as a possible carcinogen to humans [14]; radiation might cause tumors or accelerate the growth of subclinical tumors [15,16]. In recent years, head and neck injuries related to mobile phones have increased sharply [17]. A cross-sectional study [17] in the United States using a national database showed that mobile phone use can be distracting and cause injuries. In addition, increasing attention had been paid to the impact of mobile phone use on mental health (for example, addiction [18,19]), and a new term—*nomophobia*—which is short for *no mobile phone phobia* and has been considered as a symptom or syndrome of problematic digital media use in mental health [20]. However, some studies believe that the available evidence has not yet suggested that mobile phone use can cause damage to the human body (especially with respect to cancer).

Given that the use of mobile phones is growing rapidly, it is still doubted whether the improper use of mobile phones causes injuries to the human body. Our paper will provide a thorough review of literature to explore the impact of improper mobile phone use, which includes accidents and chronic disorders, on human body health.

Methods

Search Strategy

Two of the authors systematically searched PubMed, EMBASE, Cochrane, and Web of Science databases from inception to April 4, 2019. The search was limited to studies on the human body published in the English language. Additional literature was screened by manually searching the reference lists of recent reviews and studies for papers meeting the inclusion criteria.

Inclusion and Exclusion Criteria

According to the International Statistical Classification of Diseases, tenth revision [21], accidents are defined as unplanned events that sometimes have inconvenient or undesirable consequences, at times being inconsequential, and which include transport accidents and other injuries. In our study, we used the term *chronic disorders* for all nonaccident outcomes, including neoplasms (brain tumor, thyroid cancer, glioma and astrocytoma); mental health disorders such as attention-deficit/hyperactivity disorder (ADHD), nomophobia-anxiety, insecurity, anger, or discomfort; headaches; sleep disorders; injuries to the head (eye, ear, oral); injuries to the wrist; diseases of male genital organs; and other unspecific disorders including DNA damage, genotoxic effects, blood–cerebrospinal fluid barrier damage, serum S100B levels damage, total prostate specific antigen (tPSA) disorder, free prostate specific antigen (fPSA) disorder, fPSA/tPSA disorder, poor DNA integrity, chromosomal damage. The term *nomophobia*, constructed on the definitions described in the Diagnostic and Statistical Manual of Mental Disorders (Fourth edition) [22] and labeled as a “phobia for a particular/specific thing,” was used to describe the psychological condition when people had a fear of being detached from mobile phone connectivity.

Our inclusion criteria were studies that focused on (1) damage, including accidents and chronic disorders, instead of promoting healthy outcomes; (2) the use of mobile phones, including digital phone and mobile phone radio frequency radiation; (3) improper use of mobile phone, including inappropriate use occasions (eg, using mobile phone while driving or cycling), long-time or long-term use of mobile phone, and using the phone in an incorrect posture; (4) accidents occurring during mobile phone use or chronic disorders resulting from mobile phone use rather than those from any other cause (eg, occupational injuries); and studies that were (5) published in English and with (6) outcome indicators, including odd ratios (OR) or relative risk (RR) and 95% confidence intervals or mean and standard deviation.

Abstracts, comments, conferences, replies, responses, reviews (including systematic reviews), case reports, and animal studies were excluded. Additionally, studies with incomplete data and duplicate studies were also excluded.

Data Extraction and Quality Assessment

The 2 authors worked simultaneously, but independently, to screen studies, extract data from studies meeting the inclusion criteria, and assess the quality of these studies. Each author’s results were cross-checked by the other, and any disagreements on study selection, data extraction, and study quality assessment were resolved by another author.

The following information was collected using standardized data extraction forms: author information, publication year, study design, participant age, sample size, study area, measures of mobile phone use, measures of outcome-related behavior, and key outcomes.

The Newcastle-Ottawa Scale [23] was designed for the evaluation of case-control studies and cohort studies. The evaluation criteria for cross-sectional studies included 11 items recommended by the Agency for Healthcare Research and Quality [24]. The quality of each study was graded as good, fair, or poor. To be rated as good, studies needed to meet all criteria. A study was rated as poor when 1 (or more) domain was assessed as having a serious flaw. Studies that met some but not all criteria were rated as fair.

Data Analysis

A random-effects model was used to calculate overall pooled estimates. Tests for heterogeneity between studies’ results were

performed with the Cochran Q statistic and were quantified with the I^2 statistic.

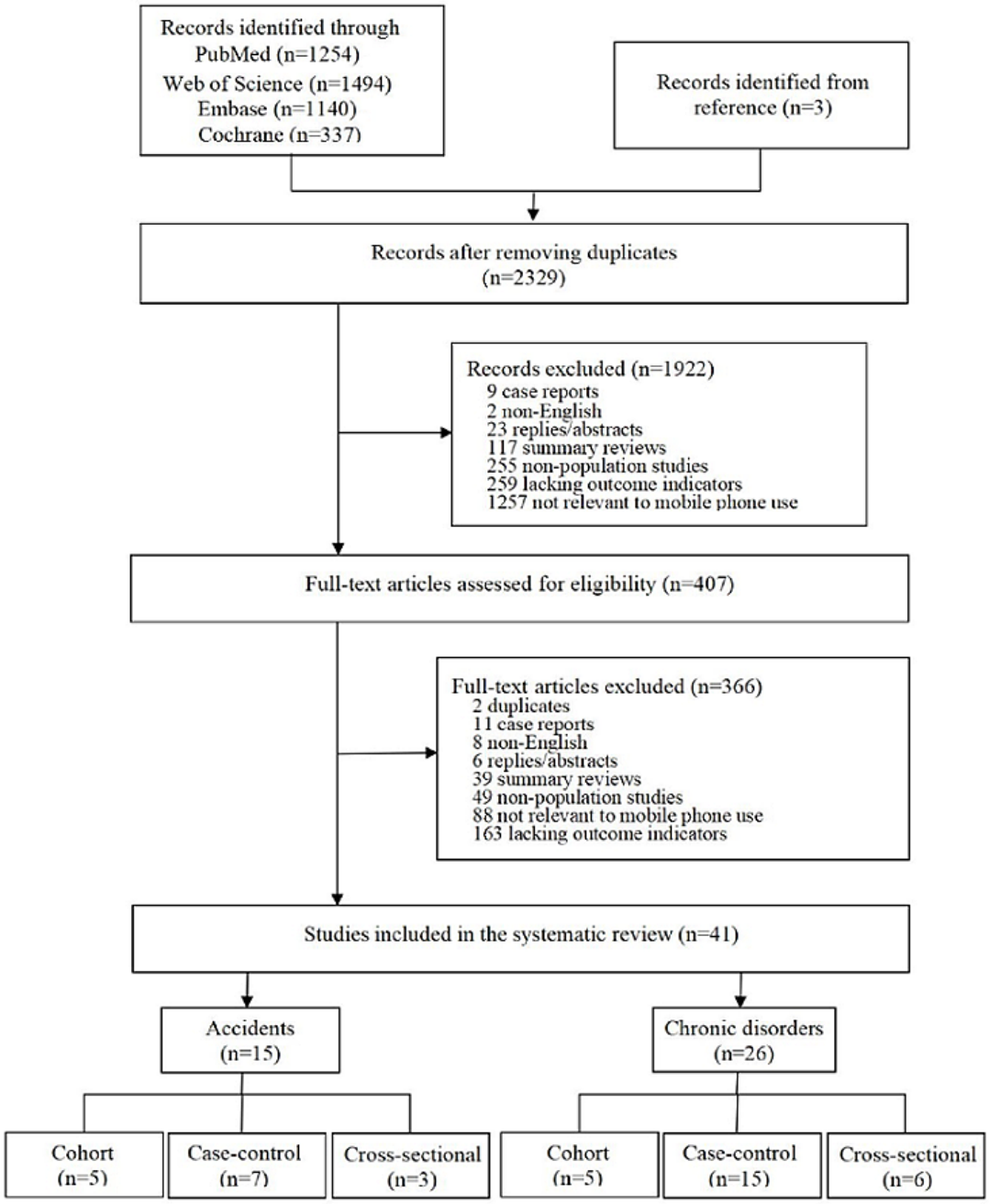
To examine the robustness of the findings, we performed subgroup analyses by country, participant age, sample size, and study-specific outcomes (accidents and chronic disorders). To validate the robustness of the findings, we performed a sensitivity analysis. The potential for publication bias was graphically explored with funnel plots, and publication bias was tested for significance with the Egger test and Begg test. All statistical procedures were 2-tailed with a significance level of 0.05 and were conducted using Stata software (version 13.0; StataCorp LLC).

Results

Study Inclusion

A total of 4228 studies were identified by the initial database search, and 3 studies were obtained by searching references; 2329 studies remained after the removal of duplicates (Figure 1). After screening titles and abstracts, 1922 records were excluded because they did not meet the selection criteria: case reports (n=9), summary reviews (n=117), nonpopulation studies (n=255), not about mobile phone use (n=1257), non-English (n=2), replies/abstracts (n=23), and no outcome indicators (n=259). Full texts of remaining papers were assessed for eligibility; 142 records were excluded because they were duplicates (n=2), case reports (n=11), summary reviews (n=39), nonpopulation research (n=49), not about mobile phone use (n=88), not English (n=8), replies/abstracts (n=6), or lacked outcome indicators (n=163). Finally, 41 studies [25-65] were included, which included cohort studies (n=10), case-control studies (n=20), and cross-sectional studies (n=11). Details are presented in Table S1 in Multimedia Appendix 1.

Figure 1. Flowchart of the selection of studies.



Study Characteristics and Quality Assessment

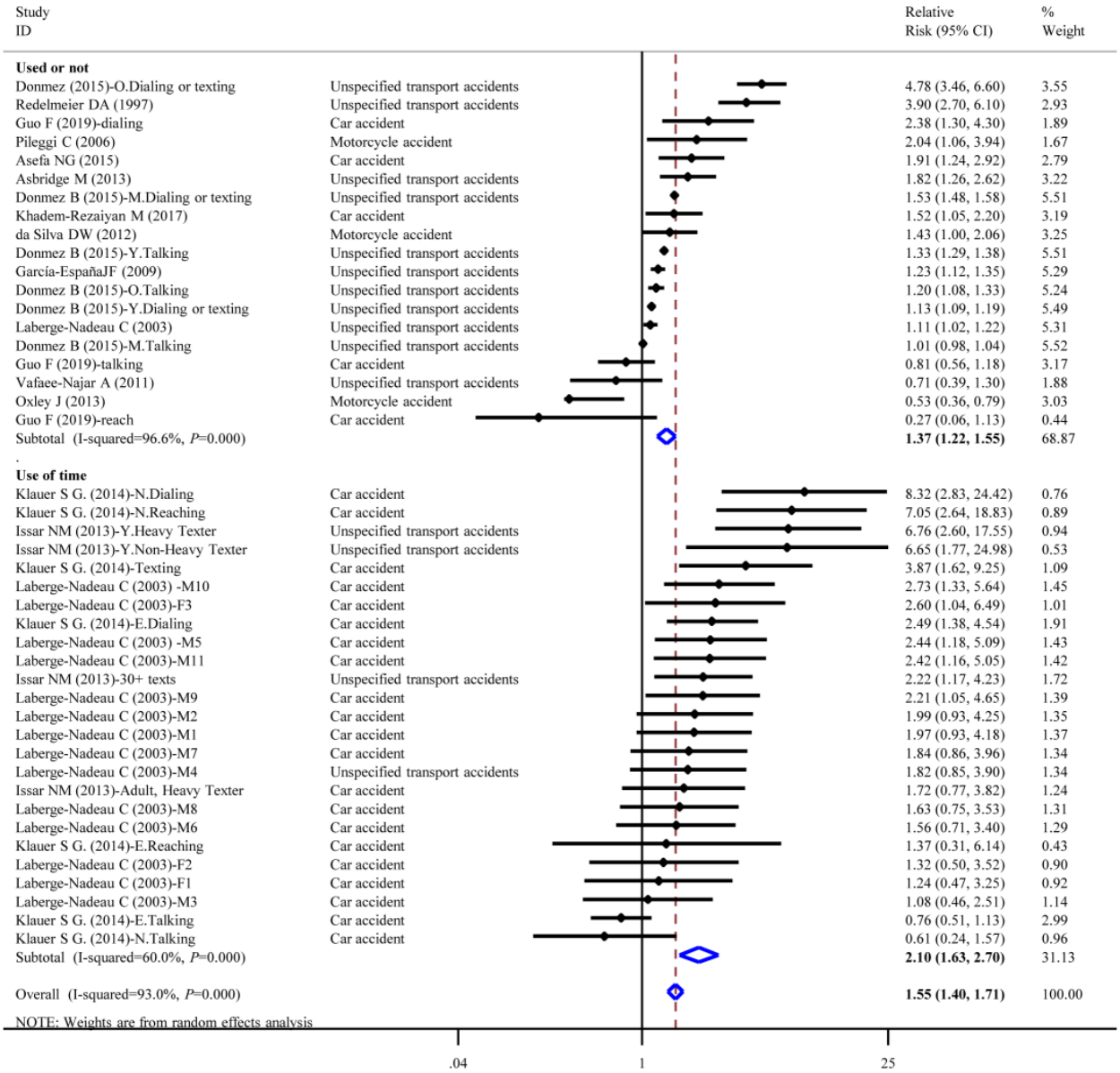
Of the 41 papers, 29 papers were published between 2011 and 2019 [25-28,30,31,35-37,41-43,45-48,50,54-65], 11 papers were published between 2002 and 2009 [32-34,38-40,44,49,51-53], and 1 paper was published in 1997 [29]. The sample sizes of the studies ranged from 6 to 15,406,515. All participants were over 7 years old. Studies were carried out in the United States (n=8) [26,30,32,37,45,50,59,64], Sweden (n=5) [38,51-53,61], Canada (n=3) [34,49,56], Korea (n=3) [27,62,63], China (n=2) [25,42], Vietnam (n=2) [31,58], Iran (n=2) [47,54], Denmark (n=1) [28], Italy (n=1) [33], Malaysia (n=1) [46], and Brazil (n=1) [57]; the remaining studies lacked relevant regional information. The outcomes were divided into accidents and chronic disorders—15 studies focused on accidents [30-34,45-47,49,50,54,56-58,64], which were mainly related to transport accidents (car accident, motorcycle accident, and unspecified transport accidents) and other accidental injuries, such as electrical injuries and explosions, and 26 studies [25-29,35-44,48,51-53,55,59-63,65] focused on chronic disorders, including neoplasms, ADHD, nomophobia, headaches, sleep disorders, dry eye diseases, ear injuries, oral problems, wrist injuries, reproductive health issues, and other unspecific chronic disorders (including DNA damage, genotoxic effects, blood-cerebrospinal fluid barrier, serum S100B levels, tPSA, fPSA, fPSA/tPSA, DNA integrity, chromosomal damage). Additional details can be found in Table S2 in Multimedia Appendix 1.

The results of the quality assessment indicated that 16 studies were good quality, and 25 were fair (Table S3 in Multimedia Appendix 1).

Mobile Phone Use and Accidents

Compared with nonmobile phone users, people who use mobile phones had a significantly higher risk for all accidents, with a pooled OR/RR of 1.55 (n=15,517,418, 95% CI 1.40 to 1.71; I^2=93.7%). The risk for mobile phone users was 1.37 times (n=15,451,501 , 95% CI 1.22 to 1.55; I^2=96.6%) that for nonmobile phone users. The top 3 relative risks were 4.78 (95% CI 3.46 to 6.60) and 3.90 (95% CI 2.70 to 6.10), both for unspecified transport accidents, and 2.38 (95% CI 1.30 to 4.30) for car accidents. Those who used mobile phones long-term had a 2.10-fold (95% CI 1.63 to 2.70) higher risk of accidents than those who did not use mobile phones or who used them for short-term; the top 3 relative risks were 8.32 (95% CI 2.83 to 24.42), 7.05 (95% CI 2.64 to 18.83), and 6.76 (95% CI 2.60 to 17.55) for car accidents, car accidents, and unspecified transport accidents, respectively (Figure 2).

Figure 2. Forest plot of accident risk and mobile phone use.

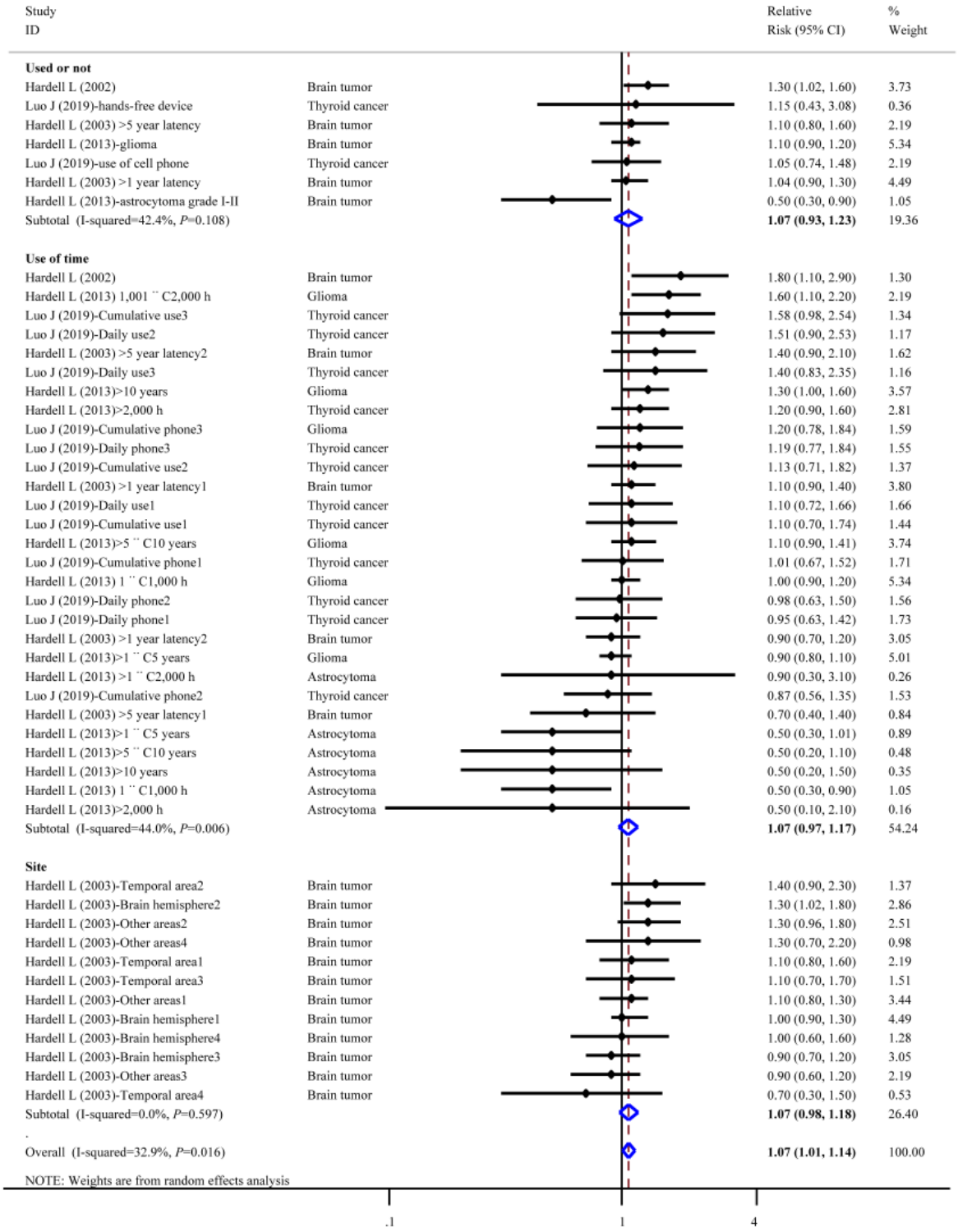


Mobile Phone Use and Chronic Disorders

The pooled risk of chronic disorders caused by mobile phone use was 1.07 times that of nonmobile phone use (95% CI 1.01 to 1.14; I^2=32.9%). Compared with nonmobile phone users, mobile phone users had a 1.07-fold risk of neoplasms (95% CI 0.93 to 1.23, I^2=42.4%). The top relative risks of neoplasms were for brain tumor (RR 1.30, 95% CI 1.02 to 1.60), followed by thyroid cancer (RR 1.15, 95% CI 0.93 to 1.23). For long-term mobile phone users there was a higher risk of neoplasms, with a pooled relative risk of 1.07 (95% CI 0.97 to 1.17). The top 3 relative risks for outcomes were brain tumor (RR 1.80, 95% CI 1.10 to 2.90), glioma (RR 1.60, 95% CI 1.10 to 2.20), and thyroid cancer (RR 1.58, 95% CI 0.98 to 2.54). Furthermore, the position when using mobile phone also increased the risk

of specific cancers; mobile phone users had a relative risk of 1.40 (95% CI 0.98 to 1.18; I^2=0.0%) for brain tumor compared with nonmobile phone users (Figure 3).

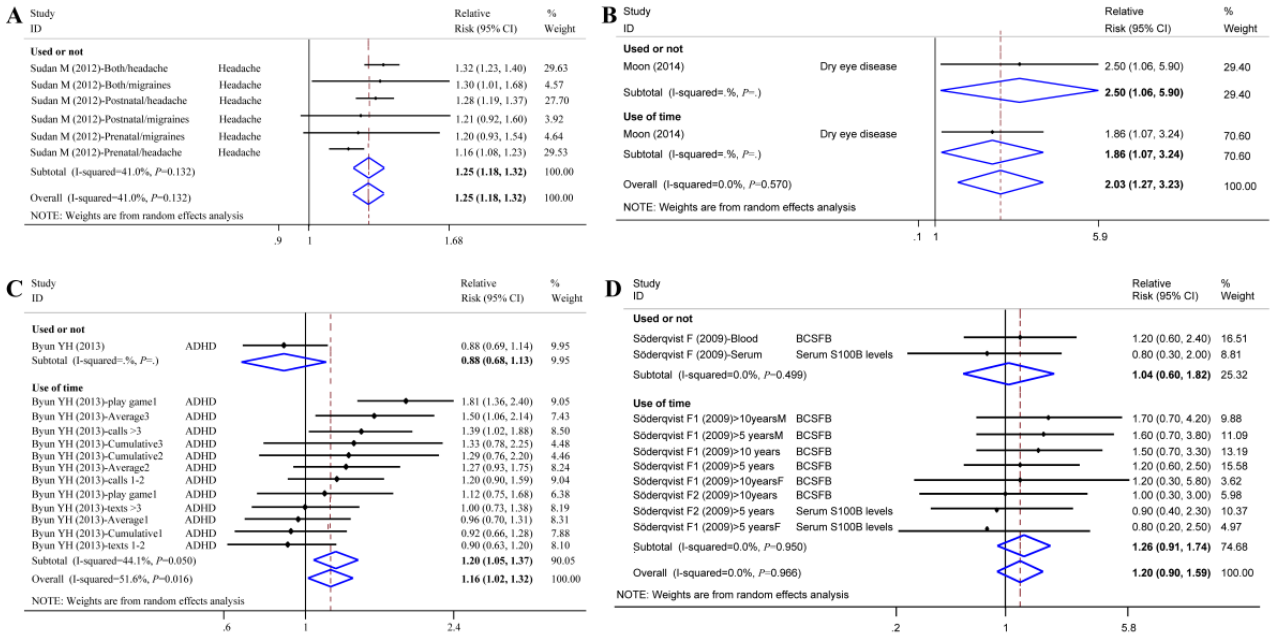
Figure 3. Forest plot of chronic disorder risk (neoplasms) and cell phone use.



Chronic nonneoplasm disorders caused by mobile phone use included mental disorders (ADHD, nomophobia), headaches, sleep disorders, injuries to the head (eye, ear, and oral), injuries to the wrist, male reproductive health issues, and other unspecific chronic disorders. Compared with nonmobile phone use, mobile phone use increased the risk of headaches (pooled risk 1.25, 95% CI 1.18 to 1.32, I^2=41.0%; Figure 4A) and the risk of dry eye disease (RR 2.03, 95% CI 1.27 to 3.23, I^2=0.0%; Figure 4B). Mobile phone users had a higher risk of ADHD than nonmobile phone users (RR 1.16, 95% CI 1.02 to 1.32, I^2=51.6%; Figure 4C), and mobile phone use increased the risk of other unspecific chronic disorders, with a pooled risk of 1.20

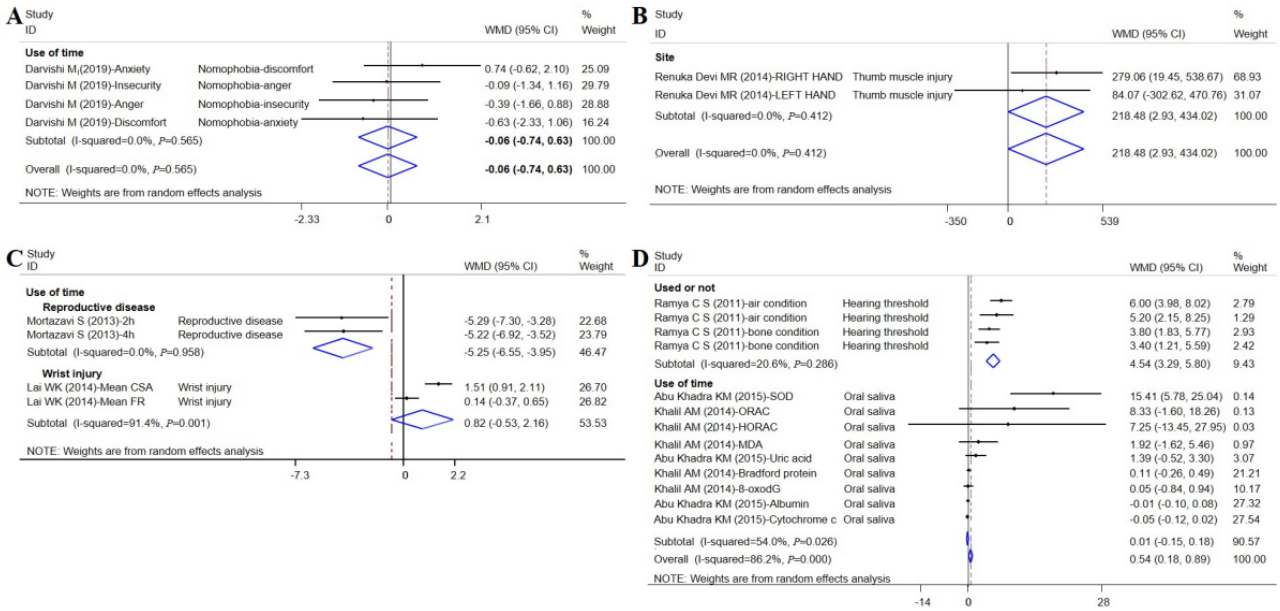
(95% CI 0.90 to 1.59, I^2=0.0%), including damage to the blood–cerebrospinal fluid barrier and elevated levels of serum S100B levels (Figure 4D).

Figure 4. Forest plot of chronic disorder risk (nonneoplasm) and cell phone use. ADHD: attention deficit/hyperactivity disorder; BCSFB: blood-cerebrospinal fluid barrier.



Compared to nonmobile phone users and short-term users, the risk for nomophobia among long-term users was –0.06 (95% CI –0.74 to 0.63; I^2=0.0%; Figure 5A); the risk was not statistically significant. Mobile phone use increased the risk of thumb injury (weighted mean difference [WMD] 218.48, 95% CI 2.93 to 434.02; I^2=0.0%; Figure 5B) and wrist extension (WMD 0.82, 95% CI –0.53 to 2.16; I^2=91.4%; Figure 5C). The risk of damage to hearing was 4.54 times higher for mobile phone users than that of the nonmobile phone users (WMD 4.54, 95% CI 3.29 to 5.80, I^2=20.6%; Figure 5D).

Figure 5. Forest plot of chronic disorder risk and cell phone use (continuous data). WMD: weight mean difference.



Subgroup Analysis

Subgroup analysis showed a consistent increase in the overall risk of cancer in the population (Table 1). Participants from the United States (OR 1.35, 95% CI 1.18 to 1.55), Denmark (OR 1.25, 95% CI 1.18 to 1.32), and aged 18 to 35 years (OR 1.62, 95% CI 1.31 to 2.00) had higher risks of injury with mobile phone use. Similarly, the larger the sample size, the higher the

risk of injury caused by the use of mobile phones. The risk of unspecified transport accidents significantly increased with mobile phone use as a result of accidents (OR 1.43, 95% CI 1.25 to 1.64). The higher risks of chronic disorders on the human body were injuries to the ear (OR 4.54, 95% CI 3.29 to 5.80), headaches (OR 1.25, 95% CI 1.18 to 1.32), and other unspecific chronic disorders (OR 0.51, 95% CI 0.04 to 0.99).

Table 1. Subgroup analyses of the risk of injuries by mobile phone use or nonuse.

Component	Studies, n (%)	Odds ratio (95% CI) (or random-effects weighted mean difference[a])
Country		
Iran	2 (2)	1.08 (0.51, 2.27)
Canada	3 (3)	1.95 (0.94, 4.07)
United States	5 (13)	1.35 (1.18, 1.55)
Denmark	1 (6)	1.25 (1.18, 1.32)
Sweden	4 (7)	1.06 (0.91, 1.24)
Sample size		
100-500	5 (8)	1.17 (0.79, 1.72)
500-1000	5 (6)	1.76 (1.14, 2.71)
>1000	10 (22)	1.21 (1.11, 1.32)
Age		
1-18 years	4 (9)	1.23 (1.15, 1.32)
18-35 years	4 (4)	1.62 (1.31, 2.00)
35-65 years	5 (7)	1.02 (0.87, 1.21)
Accidents		
Car accident	3 (5)	1.31 (0.81, 2.13)
Unspecified transport accidents	6 (11)	1.43 (1.25, 1.64)
Motorcycle accident	3 (3)	1.13 (0.51, 2.48)
Chronic disorders		
Mental disorders	2 (2)	1.37 (0.54, 3.51)
Headache	1 (6)	1.25 (1.18, 1.32)
Neoplasms	4 (7)	1.07 (0.93, 1.23)
Other unspecific chronic disorders	2 (2)	1.04 (0.60, 1.82)
Chronic disorders		
Other unspecific chronic disorders	2 (4)	0.51 (0.04, 0.99)[a]
Injuries to ear	1 (4)	4.54 (3.29, 5.80)[a]
DNA damage	1 (1)	0.13 (-0.15, 0.40)[a]

[a]Outcome measures are continuous variables; therefore, random-effects weighted mean difference was used.

Among the participants with various mobile phone use duration, Canadians and Koreans had a higher risk of injury to the human body compared with that of other populations. In studies with a participant sample size that ranged from 100 to 500 and with participants aged 18 to 35 years, there was a higher risk of accidents and chronic disorders (Table 2). In general, mobile phone use increased the risk for injury to the human body. Similarly, unspecified transport accidents were the highest cause of human body injuries as a result of accidents (OR 3.23, 95% CI 1.65 to 6.30). Increasing mobile phone use was associated with the higher risks of DNA damage (OR 7.52, 95% CI 2.23 to 12.81), male reproductive health issues (OR –4.69, 95% CI –5.64 to –3.75), and mental disorders (OR 1.20, 95% CI 1.05 to 1.37).

Table 2. Subgroup analyses of the risk of injuries by the duration of mobile phone use.

Component	Studies (included entries), n (%)	Odds ratio (95% CI) (or random-effects weighted mean difference[a])
Country		
United States	3 (23)	1.20 (0.78, 1.84)
Canada	1 (14)	1.91 (1.54, 2.35)
Korea	1 (12)	1.20 (1.05, 1.37)
Sweden	4 (37)	1.06 (0.98, 1.15)
Sample size		
100-500	4 (19)	1.89 (1.32, 2.71)
500-1000	1 (12)	1.13 (0.99, 1.28)
>1000	5 (55)	1.16 (1.07, 1.25)
Age		
1-18 years	1 (12)	1.20 (1.05, 1.37)
35-65 years	6 (41)	1.16 (1.03, 1.30)
Accidents		
Car accident	2 (21)	1.95 (1.49, 2.55)
Unspecified transport accidents	1 (4)	3.23 (1.65, 6.30)
Chronic disorders		
Mental disorders	1 (12)	1.20 (1.05, 1.37)
Tumors	4 (41)	1.07 (1.00, 1.15)
Other unspecific chronic disorders	2 (8)	1.26 (0.91, 1.74)
Chronic disorders		
Nomophobia	1 (4)	–0.06 (–0.74, 0.63)[a]
Oral problem	2 (9)	0.01 (–0.15, 0.18)[a]
DNA damage	2 (4)	7.52 (2.23, 12.81)[a]
Male reproductive health issues	1 (4)	–4.69 (–5.64, –3.75)[a]
Injuries to wrist	1 (2)	0.82 (–0.53, 2.16)[a]

[a]Outcome measures are continuous variables; therefore, random-effects weighted mean difference was used.

Publication Bias

Research results that are statistically significant may be more likely to be reported and published than results that are insignificant and invalid. In our study, the funnel plot was generally symmetric, indicating the absence of publication bias (Figure S1 in Multimedia Appendix 1).

Discussion

Principal Findings

Our review included large participant-level cohort, cross-sectional, and case-control studies on the impact of mobile phone use on outcomes related to harm to the human body. The findings suggested that mobile phone use increased the risk of accidents and chronic disorders involving the human body. Mobile phone use increased the risk of accidents by 55%. Car accidents had the highest relative risk of traffic injuries for mobile phone users. Mobile phone use also increased the risk of chronic disorders, increasing the risk of neoplasms, ADHD, headaches, and eye injuries by 7%, 16%, 25%, and 103%, respectively.

Consistent with the findings of previous studies [66-69], mobile phone use while driving increased the risk of accidents, given that it may lead to decreased situational awareness and deteriorated driving performance. Phone use while driving has become a priority road safety issues, and although it is difficult to assess the absolute increased risk for collision due to distraction of drivers caused by using mobile phones, driving simulator [6] and real-world [67] naturalistic driving studies have shown that the risk for talking on the phone while driving is significantly higher than that for undistracted driving and is comparable to the risk of driving while drunk. Ludovic et al [70] found that mobile phone use while driving was a significant distraction—even when a user is not using a mobile phone, the vibration or beeping of the phone will attract the user's attention,

thus becoming a cause of motor vehicle crashes. Drivers were more likely to miss traffic signals and were involved twice as often in car crashes when having a phone conversation while driving. In addition, visual manual tasks such as texting or typing were more likely to increase the risk of traffic accidents than other types of observable distractions [70,71]. Some interventional driving strategies and preventive measures have reduced the risk of traffic accidents among people, such as graduated driver licensing programs or advertising campaigns [72]. For example, United States, Great Britain, Canada, South Africa, and Australia have developed and use a graduated driver licensing program, which allows drivers to gain experience in low-risk driving conditions by adding an intermediate phase between the learning stage and the acquisition of the driving license [73]. Some studies [74,75] showed that the effectiveness of educational and preventive road safety programs is yet to be confirmed.

Although the risk of neoplasm from mobile phone use is still unclear, our meta-analysis suggests that improper use of mobile phones increases the risk of brain tumor, glioma, and thyroid cancer. Mobile phone radiation has been classified as possibly carcinogenic to humans [76]. There appears to be sufficient evidence that radiofrequency electromagnetic fields can cause nonthermal biological effects even when they do not cause tissue heating [77,78]. Previous evidence of damage from radiofrequency electromagnetic fields is the strongest for cancers caused by long-term exposure to mobile phones, especially brain tumor gliomas, glioblastomas, and acoustic neuromas [79,80]. In fact, the rates of brain tumors are increasing in Sweden, and the use of phones has been suggested to be the cause [81]. Little et al [82] found that ever having used mobile phones is not significantly associated with risk of glioma, but there could be increased risk for long-term users. Incorrect phone posture can increase the risk of wrist damage, chronic neck pain, and chronic shoulder pain, and the pain and fatigue worsen with longer mobile phone use [83-85]. When people use mobile phones, their body is relaxed, and their neck is prone to be bent. Hansraj et al [86] showed that there was a positive correlation between neck flexion and neck force, as well as head and neck posture in cervical spine stress and related neck pain. In addition, the long-term use of mobile phones may lead to ADHD in children and nomophobia. Studies have shown that adolescents with ADHD use electronic products significantly more often, and they usually have more sleep-wake problems [87]. Mobile phones are playing an increasingly important role in our lives. People have become dependent on mobile phones and suffer from *no mobile phonephobia* (ie, when not having a mobile phone, individuals feel discomfort, insecurity, anxiety, or anger), although the definition of *nomophobia* is not standardized, scholars have shown increasing interest and relevant scales have been designed and adjusted for different regions [88]. Finally, our meta-analysis demonstrated that mobile phone use can cause other chronic disorders, such as DNA damage (WMD 0.13, 95% CI –0.15 to 0.40). Several studies [89,90] have shown that radiofrequency radiation exposure can lead to oxidative stress in various tissues. Oxidative stress is known to play a central role in the development of cancer and aging, and it serves as a signaling agent in the inflammatory response. Recent studies [15,91] reported that the radiofrequency radiation emitted from mobile phones causes oxidative stress. Oxidative stress related to radiofrequency radiation leads to lipid, protein, and DNA damage in various tissues [15]. Our findings suggest that although the current allowable mobile phone radiation level is very low, it may be sufficient to induce biological effects. Some studies [82,92,93] have reported that existing data are not sufficient to support the assumption that tumors are caused by mobile phone usage. Thus, determination of whether these effects might cause any significant health effects requires further investigation, especially with respect to neoplasms.

Limitations

The inclusion criteria for our study were rigorous, and thus, some reports were excluded. For example, the incidences of taking selfies and sharing them on social media as well as selfie-related behaviors are increasing, particularly among young people, which possibly leads to selfie-related trauma [94,95]. Other studies [96,97,98] have reported physical harm caused by mobile phones, such as ear trauma, thigh injuries, electrical burns, and injuries caused by phone explosions. Furthermore, it has been suggested that electromagnetic fields generated by mobile phone may have long-term harmful effects, including an increase in infertility, Alzheimer disease, and other neurodegenerative diseases [99].

Our study also has some limitations. First, “damage” and “injury” were used as search queries in our study to retrieve papers on the health effects of mobile phones, other adverse outcomes caused by phone use may have been missed. Second, only 10 of the 41 studies were longitudinal studies. Additional longitudinal studies could confirm the causal relationship between mobile phone use and human health. Third, the different environments and behaviors of using mobile phones might lead to different risks of injury. We did not consider different patterns or reasons for using mobile phones in different regions and by different people, and we did not further analyze specific types and purposes of using mobile phones, such as texting or making phone calls. Finally, there was heterogeneity in our study (I^2>75%); therefore, we performed subgroup analyses to explore the source of heterogeneity.

Conclusions

There is growing evidence that mobile phone use affects the human body. Our study suggests that the use of mobile phones causes not only accidents but also chronic disorders to the human body. Although some findings are still controversial, the harm that mobile phones cause to the human body cannot be underestimated, and more research is needed to explore the direct evidence of damage to the human body.

Acknowledgments

This study was partly funded by the National Natural Science Foundation of China (grants 91746205, 71910107004, 71673199), and the China Medical Board (grant 16-262). The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The two corresponding authors had full access to all the data in the study and had final responsibility for the decision to submit the study for publication.

Authors' Contributions

XC and YC are co-first authors. PJ (jiapengff@hotmail.com) and YW (wyg@tmu.edu.cn) are co-corresponding authors.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Literature search strategy, basic characteristics of studies, quality assessment, and publication bias test.
[PDF File (Adobe PDF File), 774 KB - jmir_v24i1e21313_app1.pdf]

References

1. Mobility Reports. Ericsson. URL: https://www.ericsson.com/en/mobility-report/reports [accessed 2019-09-01]
2. Gartner. Gartner: Global smartphone sales are expected to reach 1.5 billion units in 2019, down 2.5% year-on-year. 199IT 中国互联网数据咨询网. 2019 Aug 7. URL: http://www.199it.com/archives/917684.html [accessed 2019-09-01]
3. Indraratna P, Tardo D, Yu J, Delbaere K, Brodie M, Lovell N, et al. Mobile phone technologies in the management of ischemic heart disease, heart failure, and hypertension: systematic review and meta-analysis. JMIR Mhealth Uhealth 2020 Jul 06;8(7):e16695 [FREE Full text] [doi: 10.2196/16695] [Medline: 32628615]
4. Whitehead L, Seaton P. The effectiveness of self-management mobile phone and tablet apps in long-term condition management: a systematic review. J Med Internet Res 2016 May 16;18(5):e97 [FREE Full text] [doi: 10.2196/jmir.4883] [Medline: 27185295]
5. Wang Y, Xue H, Huang Y, Huang L, Zhang D. A systematic review of application and effectiveness of mHealth interventions for obesity and diabetes treatment and self-management. Adv Nutr 2017 May;8(3):449-462 [FREE Full text] [doi: 10.3945/an.116.014100] [Medline: 28507010]
6. Wijayaratna KP, Cunningham ML, Regan MA, Jian S, Chand S, Dixit VV. Mobile phone conversation distraction: understanding differences in impact between simulator and naturalistic driving studies. Accid Anal Prev 2019 Aug;129:108-118. [doi: 10.1016/j.aap.2019.04.017] [Medline: 31150917]
7. Sun M, Sun X, Shan D. Pedestrian crash analysis with latent class clustering method. Accid Anal Prev 2019 Mar;124:50-57. [doi: 10.1016/j.aap.2018.12.016] [Medline: 30623856]
8. de Siqueira EC, de Souza FTA, Gomez RS, Gomes CC, de Souza RP. Does cell phone use increase the chances of parotid gland tumor development? a systematic review and meta-analysis. J Oral Pathol Med 2017 Aug 24;46(7):480-483. [doi: 10.1111/jop.12531] [Medline: 27935126]
9. Wall S, Wang ZM, Kendig T, Dobraca D, Lipsett M. Real-world cell phone radiofrequency electromagnetic field exposures. Environ Res 2019 Apr;171:581-592. [doi: 10.1016/j.envres.2018.09.015] [Medline: 30448205]
10. Ahlbom A, Feychting M. Mobile telephones and brain tumours. BMJ 2011 Oct 19;343:d6605. [doi: 10.1136/bmj.d6605] [Medline: 22016440]
11. De Iuliis GN, Newey RJ, King BV, Aitken RJ. Mobile phone radiation induces reactive oxygen species production and DNA damage in human spermatozoa in vitro. PLoS One 2009 Jul 31;4(7):e6446 [FREE Full text] [doi: 10.1371/journal.pone.0006446] [Medline: 19649291]
12. Stavrinos D, Byington KW, Schwebel DC. Effect of cell phone distraction on pediatric pedestrian injury risk. Pediatrics 2009 Feb;123(2):e179-e185. [doi: 10.1542/peds.2008-1382] [Medline: 19171568]
13. Cooper JM, Strayer DL. Effects of simulator practice and real-world experience on cell-phone-related driver distraction. Hum Factors 2008 Dec;50(6):893-902. [doi: 10.1518/001872008X374983] [Medline: 19292012]
14. Kim K, Kabir E, Jahan SA. The use of cell phone and insight into its potential human health impacts. Environ Monit Assess 2016 Apr 10;188(4):1-11. [doi: 10.1007/s10661-016-5227-1] [Medline: 26965900]
15. Dasdag S, Akdag MZ, Kizil G, Kizil M, Cakir DU, Yokus B. Effect of 900 MHz radio frequency radiation on beta amyloid protein, protein carbonyl, and malondialdehyde in the brain. Electromagn Biol Med 2012 Mar 23;31(1):67-74. [doi: 10.3109/15368378.2011.624654] [Medline: 22268730]
16. de Vocht F. Inferring the 1985-2014 impact of mobile phone use on selected brain cancer subtypes using Bayesian structural time series and synthetic controls. Environ Int 2016 Dec;97:100-107 [FREE Full text] [doi: 10.1016/j.envint.2016.10.019] [Medline: 27835750]

17. Povolotskiy R, Gupta N, Leverant AB, Kandinov A, Paskhover B. Head and neck injuries associated with cell phone use. JAMA Otolaryngol Head Neck Surg 2020 Feb 01;146(2):122-127 [FREE Full text] [doi: 10.1001/jamaoto.2019.3678] [Medline: 31804678]
18. Khasawneh OY. Technophobia without boarders: the influence of technophobia and emotional intelligence on technology acceptance and the moderating influence of organizational climate. Comput Human Behav 2018 Nov;88:210-218. [doi: 10.1016/j.chb.2018.07.007]
19. Xu Z, Turel O, Yuan Y. Online game addiction among adolescents: motivation and prevention factors. Eur J Inf Syst 2017 Dec 19;21(3):321-340. [doi: 10.1057/ejis.2011.56]
20. Farooqui IA, Pore P, Gothankar J. Nomophobia: an emerging issue in medical institutions? J Ment Health 2018 Oct 22;27(5):438-441. [doi: 10.1080/09638237.2017.1417564] [Medline: 29271270]
21. ICD-10 International Classification of Diseases. URL: http://icd-codes.info [accessed 2019-06-05]
22. Battle DE. Diagnostic and statistical manual of mental disorders (DSM). Codas 2013;25(2):191-192 [FREE Full text] [doi: 10.1590/s2317-17822013000200017] [Medline: 24413388]
23. Wells GA, Shea B, O'Connell D, Peterson J, Welch V, Losos M, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analysis. The Ottawa Hospital. 2004. URL: http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp [accessed 2019-09-01]
24. Methods guide for effectiveness and comparative effectiveness reviews. Agency for Healthcare Research and Quality. 2008. URL: https://effectivehealthcare.ahrq.gov/products/cer-methods-guide/overview [accessed 2021-09-13]
25. Zhang G, Yan H, Chen Q, Liu K, Ling X, Sun L, et al. Effects of cell phone use on semen parameters: Results from the MARHCS cohort study in Chongqing, China. Environ Int 2016 May;91:116-121. [doi: 10.1016/j.envint.2016.02.028] [Medline: 26949865]
26. Lewis RC, Mínguez-Alarcón L, Meeker JD, Williams PL, Mezei G, Ford JB, EARTH Study Team. Self-reported mobile phone use and semen parameters among men from a fertility clinic. Reprod Toxicol 2017 Jan;67:42-47 [FREE Full text] [doi: 10.1016/j.reprotox.2016.11.008] [Medline: 27838386]
27. Byun Y, Ha M, Kwon H, Hong Y, Leem J, Sakong J, et al. Mobile phone use, blood lead levels, and attention deficit hyperactivity symptoms in children: a longitudinal study. PLoS One 2013 Mar 21;8(3):e59742 [FREE Full text] [doi: 10.1371/journal.pone.0059742] [Medline: 23555766]
28. Sudan M, Kheifets L, Arah O, Olsen J, Zeltzer L. Prenatal and postnatal cell phone exposures and headaches in children. Open Pediatr Med Journal 2012 Dec 05;6(2012):46-52 [FREE Full text] [doi: 10.2174/1874309901206010046] [Medline: 23750182]
29. Simşek V, Sahin H, Akay AF, Kaya H, Bircan MK. The effects of cellular telephone use on serum PSA levels in men. Int Urol Nephrol 2003;35(2):193-196. [doi: 10.1023/b:urol.0000020295.92477.c4] [Medline: 15072492]
30. Issar NM, Kadakia RJ, Tsahakis JM, Yoneda ZT, Sethi MK, Mir HR, et al. The link between texting and motor vehicle collision frequency in the orthopaedic trauma population. J Inj Violence Res 2013 Jul 01;5(2):95-100 [FREE Full text] [doi: 10.5249/jivr.v5i2.330] [Medline: 23416747]
31. Truong LT, Nguyen HTT, De Gruyter C. Mobile phone use while riding a motorcycle and crashes among university students. Traffic Inj Prev 2019 Mar 19;20(2):204-210. [doi: 10.1080/15389588.2018.1546048] [Medline: 30888875]
32. García-España JF, Ginsburg KR, Durbin DR, Elliott MR, Winston FK. Primary access to vehicles increases risky teen driving behaviors and crashes: national perspective. Pediatrics 2009 Oct 28;124(4):1069-1075. [doi: 10.1542/peds.2008-3443] [Medline: 19810156]
33. Pileggi C, Bianco A, Nobile CG, Angelillo IF. Risky behaviors among motorcycling adolescents in Italy. J Pediatr 2006 Apr;148(4):527-532. [doi: 10.1016/j.jpeds.2005.11.017] [Medline: 16647418]
34. Laberge-Nadeau C, Maag U, Bellavance F, Lapierre SD, Desjardins D, Messier S, et al. Wireless telephones and the risk of road crashes. Accid Anal Prev 2003 Sep;35(5):649-660. [doi: 10.1016/s0001-4575(02)00043-x]
35. Souza LDCM, Cerqueira EDMM, Meireles JRC. Assessment of nuclear abnormalities in exfoliated cells from the oral epithelium of mobile phone users. Electromagn Biol Med 2014 Jun 28;33(2):98-102. [doi: 10.3109/15368378.2013.783856] [Medline: 23713418]
36. Khalil AM, Abu Khadra KM, Aljaberi AM, Gagaa MH, Issa HS. Assessment of oxidant/antioxidant status in saliva of cell phone users. Electromagn Biol Med 2014 Jun 19;33(2):92-97. [doi: 10.3109/15368378.2013.783855] [Medline: 23781989]
37. Luo J, Deziel NC, Huang H, Chen Y, Ni X, Ma S, et al. Cell phone use and risk of thyroid cancer: a population-based case-control study in Connecticut. Ann Epidemiol 2019 Jan;29:39-45 [FREE Full text] [doi: 10.1016/j.annepidem.2018.10.004] [Medline: 30446214]
38. Hardell L, Hallquist A, Mild KH, Carlberg M, Påhlson A, Lilja A. Cellular and cordless telephones and the risk for brain tumours. Eur J Cancer Prev 2002 Aug;11(4):377-386. [doi: 10.1097/00008469-200208000-00010] [Medline: 12195165]
39. Tiwari R, Lakshmi NK, Surender V, Rajesh ADV, Bhargava SC, Ahuja YR. Combinative exposure effect of radio frequency signals from CDMA mobile phones and aphidicolin on DNA integrity. Electromagn Biol Med 2008 Jul 07;27(4):418-425. [doi: 10.1080/15368370802473554] [Medline: 19037791]
40. Gadhia PK, Shah T, Mistry A, Pithawala M, Tamakuwala D. A preliminary study to assess possible chromosomal damage among users of digital mobile phones. Electromagn Biol Med 2009 Jul 07;22(2-3):149-159. [doi: 10.1081/jbc-120024624]

41. Çam ST, Seyhan N. Single-strand DNA breaks in human hair root cells exposed to mobile phone radiation. Int J Radiat Biol 2012 May 13;88(5):420-424. [doi: 10.3109/09553002.2012.666005] [Medline: 22348707]
42. Lai WKC, Chiu YT, Law WS. The deformation and longitudinal excursion of median nerve during digits movement and wrist extension. Man Ther 2014 Dec;19(6):608-613. [doi: 10.1016/j.math.2014.06.005] [Medline: 25024110]
43. Renuka DMRR, Saikumar P, Pradhosh. Effect of excessive mobile phone text messaging on thumb muscle. Res J Pharm Biol Chem Sci 2014;5(4):1363-1366 [FREE Full text]
44. Zeni O, Romanò M, Perrotta A, Lioi M, Barbieri R, d'Ambrosio G, et al. Evaluation of genotoxic effects in human peripheral blood leukocytes following an acute in vitro exposure to 900 MHz radiofrequency fields. Bioelectromagnetics 2005 May 14;26(4):258-265. [doi: 10.1002/bem.20078] [Medline: 15832336]
45. Donmez B, Liu Z. Associations of distraction involvement and age with driver injury severities. J Safety Res 2015 Feb;52:23-28. [doi: 10.1016/j.jsr.2014.12.001] [Medline: 25662879]
46. Oxley J, Yuen J, Ravi MD, Hoareau E, Mohammed MAA, Bakar H, et al. Commuter motorcycle crashes in Malaysia: an understanding of contributing factors. Ann Adv Automot Med 2013;57:45-54 [FREE Full text] [Medline: 24406945]
47. Vafaeenajar A, Khabbazkhoob M, Alidadisoltangholi H. Investigating the relative risk factors of injuries caused by accidents on roads in the Mashhad area in 2007. Iran Red Crescent Med J 2011 Aug 1;13(8):530-536 [FREE Full text]
48. Mortazavi S, Parsanezhad M, Kazempour M, Ghahramani P, Mortazavi A, Davari M. Male reproductive health under threat: Short term exposure to radiofrequency radiations emitted by common mobile jammers. J Hum Reprod Sci 2013 Apr;6(2):124-128 [FREE Full text] [doi: 10.4103/0974-1208.117178] [Medline: 24082653]
49. Redelmeier DA, Tibshirani RJ. Association between cellular-telephone calls and motor vehicle collisions. N Engl J Med 1997 Feb 13;336(7):453-458. [doi: 10.1056/NEJM199702133360701] [Medline: 9017937]
50. Guo F, Kim I, Klauer SG. Semiparametric Bayesian models for evaluating time-variant driving risk factors using naturalistic driving data and case-crossover approach. Stat Med 2019 Jan 30;38(2):160-174. [doi: 10.1002/sim.7574] [Medline: 29280183]
51. Hardell L, Mild K, Carlberg M. Further aspects on cellular and cordless telephones and brain tumours. Int J Oncol 2003 Feb 01:399-407. [doi: 10.3892/ijo.22.2.399] [Medline: 12527940]
52. Söderqvist F, Carlberg M, Hardell L. Mobile and cordless telephones, serum transthyretin and the blood-cerebrospinal fluid barrier: a cross-sectional study. Environ Health 2009 Apr 21;8:19 [FREE Full text] [doi: 10.1186/1476-069X-8-19] [Medline: 19383125]
53. Söderqvist F, Carlberg M, Hardell L. Use of wireless telephones and serum S100B levels: a descriptive cross-sectional study among healthy Swedish adults aged 18-65 years. Sci Total Environ 2009 Jan 01;407(2):798-805. [doi: 10.1016/j.scitotenv.2008.09.051] [Medline: 18986685]
54. Khadem-Rezaiyan M, Moallem SR, Vakili V. High-risk behaviors while driving: a population-based study from Iran. Traffic Inj Prev 2017 Apr 03;18(3):257-261. [doi: 10.1080/15389588.2016.1192612] [Medline: 27260770]
55. Darvishi M, Noori M, Nazer MR, Sheikholeslami S, Karimi E. Investigating different dimensions of nomophobia among medical students: a cross-sectional study. Open Access Maced J Med Sci 2019 Feb 28;7(4):573-578 [FREE Full text] [doi: 10.3889/oamjms.2019.138] [Medline: 30894914]
56. Asbridge M, Brubacher JR, Chan H. Cell phone use and traffic crash risk: a culpability analysis. Int J Epidemiol 2013 Feb 18;42(1):259-267. [doi: 10.1093/ije/dys180] [Medline: 23159829]
57. da Silva DW, de Andrade SM, Soares DFPDP, Mathias TADF, Matsuo T, de Souza RKT. Factors associated with road accidents among Brazilian motorcycle couriers. ScientificWorldJournal 2012;2012:605480 [FREE Full text] [doi: 10.1100/2012/605480] [Medline: 22629158]
58. Asefa NG, Ingale L, Shumey A, Yang H. Prevalence and factors associated with road traffic crash among taxi drivers in Mekelle town, northern Ethiopia, 2014: a cross sectional study. PLoS One 2015 Mar 17;10(3):e0118675 [FREE Full text] [doi: 10.1371/journal.pone.0118675] [Medline: 25781940]
59. Fuller C, Lehman E, Hicks S, Novick MB. Bedtime use of technology and associated sleep problems in children. Glob Pediatr Health 2017;4:2333794X17736972 [FREE Full text] [doi: 10.1177/2333794X17736972] [Medline: 29119131]
60. Abu Khadra KM, Khalil AM, Abu Samak M, Aljaberi A. Evaluation of selected biochemical parameters in the saliva of young males using mobile phones. Electromagn Biol Med 2015 Mar 05;34(1):72-76. [doi: 10.3109/15368378.2014.881370] [Medline: 24499288]
61. Hardell L, Carlberg M. Use of mobile and cordless phones and survival of patients with glioma. Neuroepidemiology 2013;40(2):101-108. [doi: 10.1159/000341905] [Medline: 23095687]
62. Mortazavi SMJ, Mortazavi SAR, Paknahad M. Association between exposure to smartphones and ocular health in adolescents. Ophthalmic Epidemiol 2016 Dec 23;23(6):418-418. [doi: 10.1080/09286586.2016.1212992] [Medline: 27552364]
63. Moon JH, Lee MY, Moon NJ. Association between video display terminal use and dry eye disease in school children. J Pediatr Ophthalmol Strabismus 2014 Feb 04;51(2):87-92. [doi: 10.3928/01913913-20140128-01] [Medline: 24495620]
64. Klauer SG, Guo F, Simons-Morton BG, Ouimet MC, Lee SE, Dingus TA. Distracted driving and risk of road crashes among novice and experienced drivers. N Engl J Med 2014 Jan 02;370(1):54-59. [doi: 10.1056/nejmsa1204142]
65. Ramya C, Vinutha S, Karthiyanee K. Effect of mobile phone usage on hearing threshold: a pilot study. Indian J Otol 2011;17(4):159. [doi: 10.4103/0971-7749.94494]

66. Al-Jasser F, Mohamed A, Choudry A, Youssef R. Mobile phone use while driving and the risk of collision: a study among preparatory year students at King Saud University, Riyadh, Saudi Arabia. J Family Community Med 2018;25(2):102-107 [FREE Full text] [doi: 10.4103/jfcm.JFCM_139_17] [Medline: 29922110]
67. Alghnam S, Towhari J, Alkelya M, Alsaif A, Alrowaily M, Alrabeeah F, et al. The association between mobile phone use and severe traffic injuries: a case-control study from Saudi Arabia. Int J Environ Res Public Health 2019 Jul 29;16(15):2706 [FREE Full text] [doi: 10.3390/ijerph16152706] [Medline: 31362446]
68. Gariazzo C, Stafoggia M, Bruzzone S, Pelliccioni A, Forastiere F. Association between mobile phone traffic volume and road crash fatalities: a population-based case-crossover study. Accid Anal Prev 2018 Jun;115:25-33. [doi: 10.1016/j.aap.2018.03.008] [Medline: 29544134]
69. Dingus TA, Guo F, Lee S, Antin JF, Perez M, Buchanan-King M, et al. Driver crash risk factors and prevalence evaluation using naturalistic driving data. Proc Natl Acad Sci U S A 2016 Mar 08;113(10):2636-2641 [FREE Full text] [doi: 10.1073/pnas.1513271113] [Medline: 26903657]
70. Gicquel L, Ordonneau P, Blot E, Toillon C, Ingrand P, Romo L. Description of various factors contributing to traffic accidents in youth and measures proposed to alleviate recurrence. Front Psychiatry 2017 Jun 01;8:94 [FREE Full text] [doi: 10.3389/fpsyt.2017.00094] [Medline: 28620324]
71. Young R. Removing biases from crash odds ratio estimates of secondary tasks: a new analysis of the SHRP 2 naturalistic driving study data. In: SAE Technical Paper Series. 2017 Presented at: WCX™ 17: SAE World Congress Experience; April 4-6; Detroit, Michigan. [doi: 10.4271/2017-01-1380]
72. Scott-Parker B, Oviedo-Trespalacios O. Young driver risky behaviour and predictors of crash risk in Australia, New Zealand and Colombia: same but different? Accid Anal Prev 2017 Feb;99(Pt A):30-38. [doi: 10.1016/j.aap.2016.11.001] [Medline: 27865138]
73. Huemer AK, Schumacher M, Mennecke M, Vollrath M. Systematic review of observational studies on secondary task engagement while driving. Accid Anal Prev 2018 Oct;119:225-236. [doi: 10.1016/j.aap.2018.07.017] [Medline: 30055511]
74. Carcaillon LI, Salmi LR, Atout-Route Evaluation Group. Evaluation of a program to reduce motor-vehicle collisions among young adults in the county of Landes, France. Accid Anal Prev 2005 Nov;37(6):1049-1055. [doi: 10.1016/j.aap.2005.06.003] [Medline: 16036209]
75. Olsson B, Pütz H, Reitzug F, Humphreys DK. Evaluating the impact of penalising the use of mobile phones while driving on road traffic fatalities, serious injuries and mobile phone use: a systematic review. Inj Prev 2020 Aug;26(4):378-385. [doi: 10.1136/injuryprev-2019-043619] [Medline: 32229534]
76. IARC classifies radiofrequency electromagnetic fields as possibly carcinogenic to humans. International Agency for Research on Cancer. URL: http://www.iarc.fr/en/media-centre/pr/2011/pdfs/pr208_E.pdf [accessed 2021-09-13]
77. Narayanan SN, Jetti R, Kesari KK, Kumar RS, Nayak SB, Bhat PG. Radiofrequency electromagnetic radiation-induced behavioral changes and their possible basis. Environ Sci Pollut Res Int 2019 Oct;26(30):30693-30710. [doi: 10.1007/s11356-019-06278-5] [Medline: 31463749]
78. Kesari KK, Siddiqui MH, Meena R, Verma HN, Kumar S. Cell phone radiation exposure on brain and associated biological systems. Indian J Exp Biol 2013 Mar;51(3):187-200. [Medline: 23678539]
79. Prasad M, Kathuria P, Nair P, Kumar A, Prasad K. Mobile phone use and risk of brain tumours: a systematic review of association between study quality, source of funding, and research outcomes. Neurol Sci 2017 May;38(5):797-810. [doi: 10.1007/s10072-017-2850-8] [Medline: 28213724]
80. Yang M, Guo W, Yang C, Tang J, Huang Q, Feng S, et al. Mobile phone use and glioma risk: a systematic review and meta-analysis. PLoS One 2017 May 4;12(5):e0175136 [FREE Full text] [doi: 10.1371/journal.pone.0175136] [Medline: 28472042]
81. Hardell L, Carlberg M. Mobile phones, cordless phones and rates of brain tumors in different age groups in the Swedish National Inpatient Register and the Swedish Cancer Register during 1998-2015. PLoS One 2017 Oct 4;12(10):e0185461 [FREE Full text] [doi: 10.1371/journal.pone.0185461] [Medline: 28976991]
82. Little MP, Rajaraman P, Curtis RE, Devesa SS, Inskip PD, Check DP, et al. Mobile phone use and glioma risk: comparison of epidemiological study results with incidence trends in the United States. BMJ 2012 Mar 08;344:e1147 [FREE Full text] [doi: 10.1136/bmj.e1147] [Medline: 22403263]
83. Lee S, Kang H, Shin G. Head flexion angle while using a smartphone. Ergonomics 2015 Oct 17;58(2):220-226. [doi: 10.1080/00140139.2014.967311] [Medline: 25323467]
84. Xie YF, Szeto G, Madeleine P, Tsang S. Spinal kinematics during smartphone texting - a comparison between young adults with and without chronic neck-shoulder pain. Appl Ergon 2018 Apr;68:160-168. [doi: 10.1016/j.apergo.2017.10.018] [Medline: 29409630]
85. Kim S, Koo S. Effect of duration of smartphone use on muscle fatigue and pain caused by forward head posture in adults. J Phys Ther Sci 2016 Jun;28(6):1669-1672 [FREE Full text] [doi: 10.1589/jpts.28.1669] [Medline: 27390391]
86. Hansraj KK. Assessment of stresses in the cervical spine caused by posture and position of the head. Surg Technol Int 2014 Nov;25:277-279. [Medline: 25393825]

87. Sangwaiya A, Sharma J, Kaira V, Ratan K, Sen R, Gupta V. Burkitt's lymphoma masquerading as intestinal obstruction: An uncommon entity with variable clinical presentation. Clin Cancer Investig J 2014;3(5):441. [doi: 10.4103/2278-0513.138090]
88. Mireku MO, Barker MM, Mutz J, Dumontheil I, Thomas MS, Röösli M, et al. Night-time screen-based media device use and adolescents' sleep and health-related quality of life. Environ Int 2019 Mar;124:66-78 [FREE Full text] [doi: 10.1016/j.envint.2018.11.069] [Medline: 30640131]
89. Gulati S, Kosik P, Durdik M, Skorvaga M, Jakl L, Markova E, et al. Effects of different mobile phone UMTS signals on DNA, apoptosis and oxidative stress in human lymphocytes. Environ Pollut 2020 Dec;267:115632. [doi: 10.1016/j.envpol.2020.115632] [Medline: 33254645]
90. Panagopoulos DJ, Karabarbounis A, Yakymenko I, Chrousos GP. Human-made electromagnetic fields: Ion forced-oscillation and voltage-gated ion channel dysfunction, oxidative stress and DNA damage (Review). Int J Oncol 2021 Nov;59(5):92 [FREE Full text] [doi: 10.3892/ijo.2021.5272] [Medline: 34617575]
91. Dasdag S, Akdag MZ. The link between radiofrequencies emitted from wireless technologies and oxidative stress. J Chem Neuroanat 2016 Sep;75(Pt B):85-93. [doi: 10.1016/j.jchemneu.2015.09.001] [Medline: 26371078]
92. Röösli M, Lagorio S, Schoemaker MJ, Schüz J, Feychting M. Brain and salivary gland tumors and mobile phone use: evaluating the evidence from various epidemiological study designs. Annu Rev Public Health 2019 Apr 01;40(1):221-238. [doi: 10.1146/annurev-publhealth-040218-044037] [Medline: 30633716]
93. Smeds H, Wales J, Mathiesen T, Talbäck M, Feychting M. Occurrence of primary brain tumors in cochlear implant patients in Sweden between 1989 and 2014. Clin Epidemiol 2018;10:1401-1405 [FREE Full text] [doi: 10.2147/CLEP.S164556] [Medline: 30323683]
94. Jain MJ, Mavani KJ. A comprehensive study of worldwide selfie-related accidental mortality: a growing problem of the modern society. Int J Inj Contr Saf Promot 2017 Dec 14;24(4):544-549. [doi: 10.1080/17457300.2016.1278240] [Medline: 28632035]
95. Dokur M, Petekkaya E, Karadağ M. Media-based clinical research on selfie-related injuries and deaths. Ulus Travma Acil Cerrahi Derg 2018 Mar;24(2):129-135 [FREE Full text] [doi: 10.5505/tjtes.2017.83103] [Medline: 29569684]
96. Amernik K, Kabacińska A, Tarnowska C, Paradowska-Opałka B. Ostry uraz akustyczny i termiczny ucha spowodowany awarią telefonu komórkowego. Otolaryngologia Polska 2007 Jan;61(4):484-486. [doi: 10.1016/S0030-6657(07)70466-3]
97. Görgülü T, Torun M, Olgun A. A cause of severe thigh injury: battery explosion. Ann Med Surg (Lond) 2016 Feb;5:49-51 [FREE Full text] [doi: 10.1016/j.amsu.2015.12.048] [Medline: 26862395]
98. Cherubino M, Pellegatta I, Sallam D, Pulerà E, Valdatta L. Enzymatic debridement after mobile phone explosion: a case report. Ann Burns Fire Disasters 2016 Dec 31;29(4):273-275 [FREE Full text] [Medline: 28289361]
99. Bouji M, Lecomte A, Gamez C, Blazy K, Villégier A. Impact of cerebral radiofrequency exposures on oxidative stress and corticosterone in a rat model of Alzheimer's Disease. J Alzheimers Dis 2020 Jan 21;73(2):467-476. [doi: 10.3233/jad-190593]

Abbreviations

ADHD: attention-deficit/hyperactivity disorder
fPSA: free prostate specific antigen
OR: odds ratio
RR: relative risk
tPSA: total prostate specific antigen
WMD: weighted mean difference

Edited by R Kukafka, G Eysenbach; submitted 10.06.20; peer-reviewed by J Sun, H Yasuda, M Peden; comments to author 06.07.20; revised version received 05.08.20; accepted 02.08.21; published 20.01.22.

Please cite as:
Cao X, Cheng Y, Xu C, Hou Y, Yang H, Li S, Gao Y, Jia P, Wang Y
Risk of Accidents or Chronic Disorders From Improper Use of Mobile Phones: A Systematic Review and Meta-analysis
J Med Internet Res 2022;24(1):e21313
URL: https://www.jmir.org/2022/1/e21313
doi:10.2196/21313
PMID:35049511

Review

Web Portals for Patients With Chronic Diseases: Scoping Review of the Functional Features and Theoretical Frameworks of Telerehabilitation Platforms

Yuh Morimoto[1], PhD; Tetsuya Takahashi[2,3], PT, MSc, PhD; Ryuichi Sawa[2], PT, MSc, PhD; Masakazu Saitoh[2], PT, MSc, PhD; Tomoyuki Morisawa[2], PT, MSc, PhD; Nobuyuki Kagiyama[3,4], MD, PhD; Takatoshi Kasai[3,4], MD, PhD; Birthe Dinesen[5], MSc, PhD; Malene Hollingdal[6], MD, PhD; Jens Refsgaard[6], MD, PhD; Hiroyuki Daida[1,2,3,4], MD, PhD

[1]Faculty of Health Science, Juntendo University, Tokyo, Japan

[2]Department of Physical Therapy, Faculty of Health Science, Juntendo University, Tokyo, Japan

[3]Department of Digital Health and Telemedicine Research and Development, Faculty of Health Science, Juntendo University, Tokyo, Japan

[4]Department of Cardiovascular Biology and Medicine, Juntendo University Graduate School of Medicine, Tokyo, Japan

[5]Laboratory for Welfare Technologies - Telehealth & Telerehabilitation, Sport Sciences - Performance and Technology, Department of Health Science and Technology, Aalborg University, Aalborg, Denmark

[6]Department of Cardiology, Regional Hospital in Viborg, Viborg, Denmark

Corresponding Author:
Hiroyuki Daida, MD, PhD
Faculty of Health Science
Juntendo University
2-1-1 Hongo Bunkyo-ku
Tokyo, 1138421
Japan
Phone: 81 338133111 ext 3900
Fax: 81 338121781
Email: daida@juntendo.ac.jp

Abstract

Background: The COVID-19 pandemic has required an increased need for rehabilitation activities applicable to patients with chronic diseases. Telerehabilitation has several advantages, including reducing clinic visits by patients vulnerable to infectious diseases. Digital platforms are often used to assist rehabilitation services for patients in remote settings. Although web portals for medical use have existed for years, the technology in telerehabilitation remains a novel method.

Objective: This scoping review investigated the functional features and theoretical approaches of web portals developed for telerehabilitation in patients with chronic diseases.

Methods: PubMed and Web of Science were reviewed to identify articles associated with telerehabilitation. Of the 477 nonduplicate articles reviewed, 35 involving 14 portals were retrieved for the scoping review. The functional features, targeted diseases, and theoretical approaches of these portals were studied.

Results: The 14 portals targeted patients with chronic obstructive pulmonary disease, cardiovascular, osteoarthritis, multiple sclerosis, cystic fibrosis diseases, and stroke and breast cancer survivors. Monitoring/data tracking and communication functions were the most common, followed by exercise instructions and diary/self-report features. Several theoretical approaches, behavior change techniques, and motivational techniques were found to be utilized.

Conclusions: The web portals could unify and display multiple types of data and effectively provide various types of information. Asynchronous correspondence was more favorable than synchronous, real-time interactions. Data acquisition often required assistance from other digital tools. Various functions with patient-centered principles, behavior change strategies, and motivational techniques were observed for better support shifting to a healthier lifestyle. These findings suggested that web portals for telerehabilitation not only provided entrance into rehabilitation programs but also reinforced participant-centered treatment, adherence to rehabilitation, and lifestyle changes over time.

(J Med Internet Res 2022;24(1):e27759) doi:10.2196/27759

KEYWORDS

telerehabilitation; web portal; chronic disease; monitoring/data tracking function; patient-centered care

Introduction

Chronic diseases are the leading causes of death worldwide. The World Health Organization has reported that chronic diseases are responsible for almost 60% of deaths worldwide [1]. Studies show evidence of the effectiveness of rehabilitation for patients with chronic disease; however, rehabilitation programs are generally underused [2,3]. The major barriers to participation in a rehabilitation program include inconvenient timing, travel and transport issues, and lack of perceived benefit [4]. Currently, rehabilitation services are facing even more difficulty because of the COVID-19 pandemic. Conventional clinic-based rehabilitation programs have been suspended as a result of physical distancing recommendations and a shortage of health care services [5,6]. Patients with chronic diseases are more likely to develop severe conditions than patients without chronic diseases, thus visiting clinics is considered to be a risk to be avoided by patients [7]. Telerehabilitation can address such obstacles, even during the COVID-19 pandemic.

Telerehabilitation is defined as rehabilitation activities performed using information provided by communication technologies over a distance [8]. Telerehabilitation for patients with chronic diseases may consist of many components, for example, exercise instruction, education, better communication, and self-management training [9-11]. Digital platforms are often used to assist rehabilitation services for patients in remote settings [12-14], and web portals are one of the potential technologies. The term “portal (computing)” has been described as a website that is used as a gateway to the internet, where information useful to a person interested in particular topics has been gathered [15]. A web portal in clinical use has been described as a kind of electronic health record that permits patients to access their records or communicate with their health care professionals [16,17]. Furthermore, the US government describes a web portal as “a secure online website that gives patients convenient, 24-hour access to personal health information from anywhere with an internet connection” [18]. Taking into account the personal nature of the information, access is often limited to authorized people in a secure and confidential setting [19]. To encourage patient engagement and provide benefits, web portals are recommended to follow the principle of patient-centered care [20]. The primary concept of patient-centered care is that patient values guide clinical decisions [21]. The 6 factors defined as components of patient-centered care are (1) respect for patients’ values, preferences, and expressed needs; (2) coordination and integration of care; (3) information, communication, and education; (4) physical comfort; (5) emotional support; and (6) involvement of family and friends [22].

Web portals have the potential to be a core digital component for patient-centered telerehabilitation, however, relatively few have been adapted for telerehabilitation [23], and few studies to date have analyzed the use of web portals for telerehabilitation in an international context. It is not known whether there are unique characteristics in web portals developed for telerehabilitation targeting chronic diseases rather portals designed for other medical purposes. For these reasons, a scoping review was conducted to gather knowledge about what has been designed and tested in clinical practice, as well as to identify any existing gaps in knowledge. The search strategy according to population, concept, and context elements was as follows: P, telerehabilitation participants with chronic disease; C, web portals developed for telerehabilitation, functional features, theoretical approaches, behavior change techniques (BCT), motivational techniques (MTs), and mode of delivery; C, any gender, age, or region. The following research question was formulated to guide this review: What are the functional features and theoretical approaches of web portals developed for telerehabilitation programs in patients with chronic diseases, as well as any characteristics that can be observed through the investigations?

Methods

Study Design

This study was designed to map the functional features of web portals utilized for telerehabilitation targeting chronic diseases. Although systematic reviews are preferred for answering clearly defined questions, a scoping review method is considered useful for answering broad questions [24]. The scoping review protocol in this study was developed using the guidelines provided by the Preferred Reporting Items for Systematic Reviews and Meta-analysis Extension for Scoping Reviews (PRISMA-ScR) checklist [24,25]. With the cooperation of an experienced university librarian, multiple databases were searched, ensuring the appropriateness of the search strategy. The final version of the protocol is available from a protocol registry [26].

Identification and Selection of the Relevant Articles

A literature search of the international online bibliographic databases PubMed [27] and the Web of Science [28] was performed on May 17, 2021. The search formula used for the PubMed search engine was ((telerehabilitation) AND (chronic)) OR ((telerehabilitation) AND ((portal) OR (web-based) OR (digital platform) OR (online platform) OR (internet-based))). Because the definitions of web portals for telerehabilitation participants were not determined in a common context, digital platforms used for telerehabilitation were reviewed with the principal concepts of web portals described above, such as (1) web-based application, (2) allows the participants 24-hour access, and (3) containing useful rehabilitation-related functions for participants. To extract core components of web portals for telerehabilitation, portals developed for purposes other than telerehabilitation were excluded, as were clinician portals. To focus on the essential features, telerehabilitation portals targeting anything other than chronic diseases were also excluded. Web applications designed for a single purpose, web services with

limited time of access, and digital platforms based on native applications (eg, iOS and Android apps and PC applications) were excluded based on the definitions of web portals described above. Abstracts of the selected studies were reviewed by the team members according to the inclusion/exclusion criteria. Review articles, editorials, conference reports, and interviews were also excluded, whereas protocols were included.

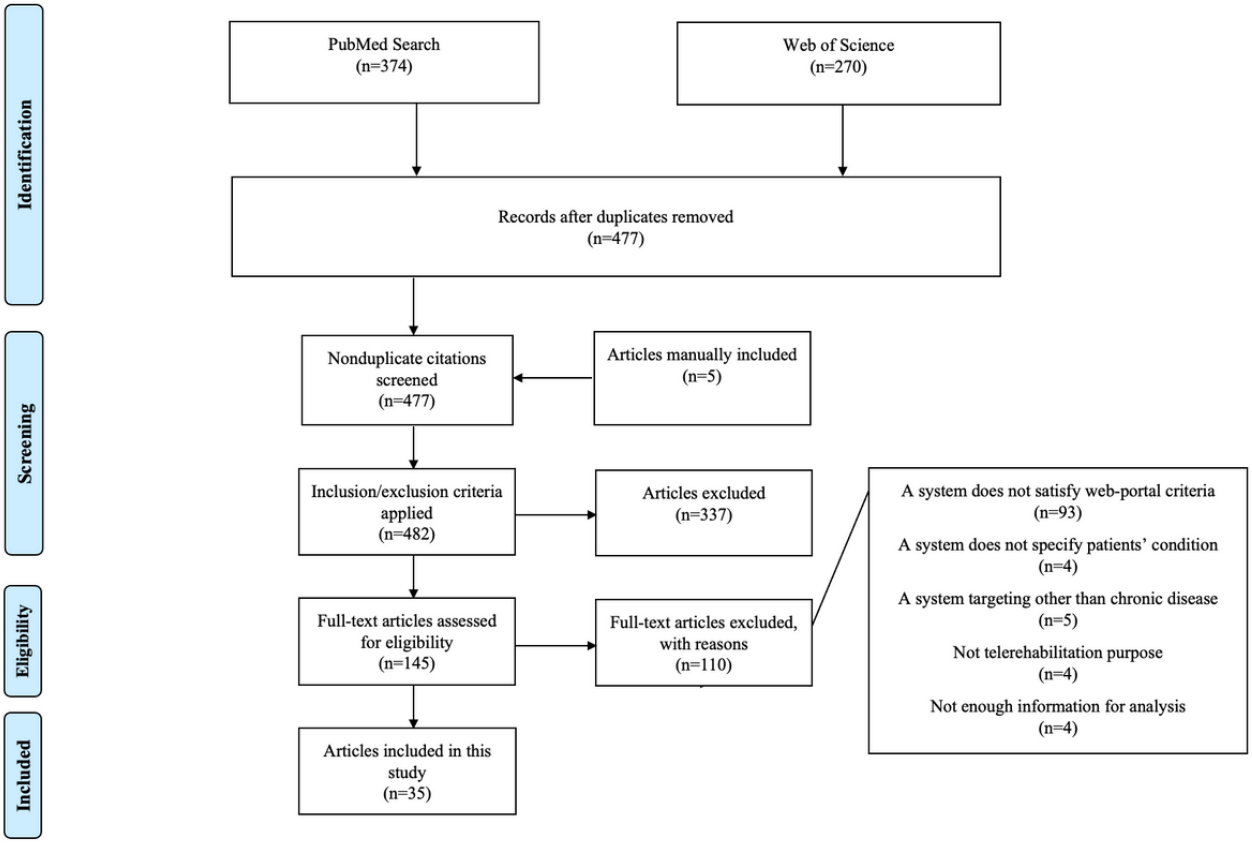
Data extraction and charting were performed according to Arksey and O'Malley's guideline [29]. Data from the selected articles were extracted with the following modules: portal name, project name, country, functional features, targeted chronic diseases, language, and other systems required for the telerehabilitation service. Theories and models used in a study were investigated as well as BCTs, MTs, and mode of delivery. Several articles were found to describe the same telerehabilitation project. The names of the projects and web portals, as well as the authors, institutions, and trial registration numbers, were used to identify telerehabilitation projects described in multiple articles.

Results

Overview

Of the 644 articles initially identified by keywords, 477 remained after the removal of duplicates. An additional 5 articles were found through manual searching. Figure 1 shows the flow diagram of the article identification process. The titles and abstracts of these 482 nonduplicated articles were reviewed during the initial screening phase, resulting in the identification of 145 articles for full-text screening. Of these 145 articles, 35 articles with 14 platforms satisfied the criteria and were included in the analysis. Full reasons for exclusion are shown in Figure 1.

Figure 1. Flow diagram of article identification.

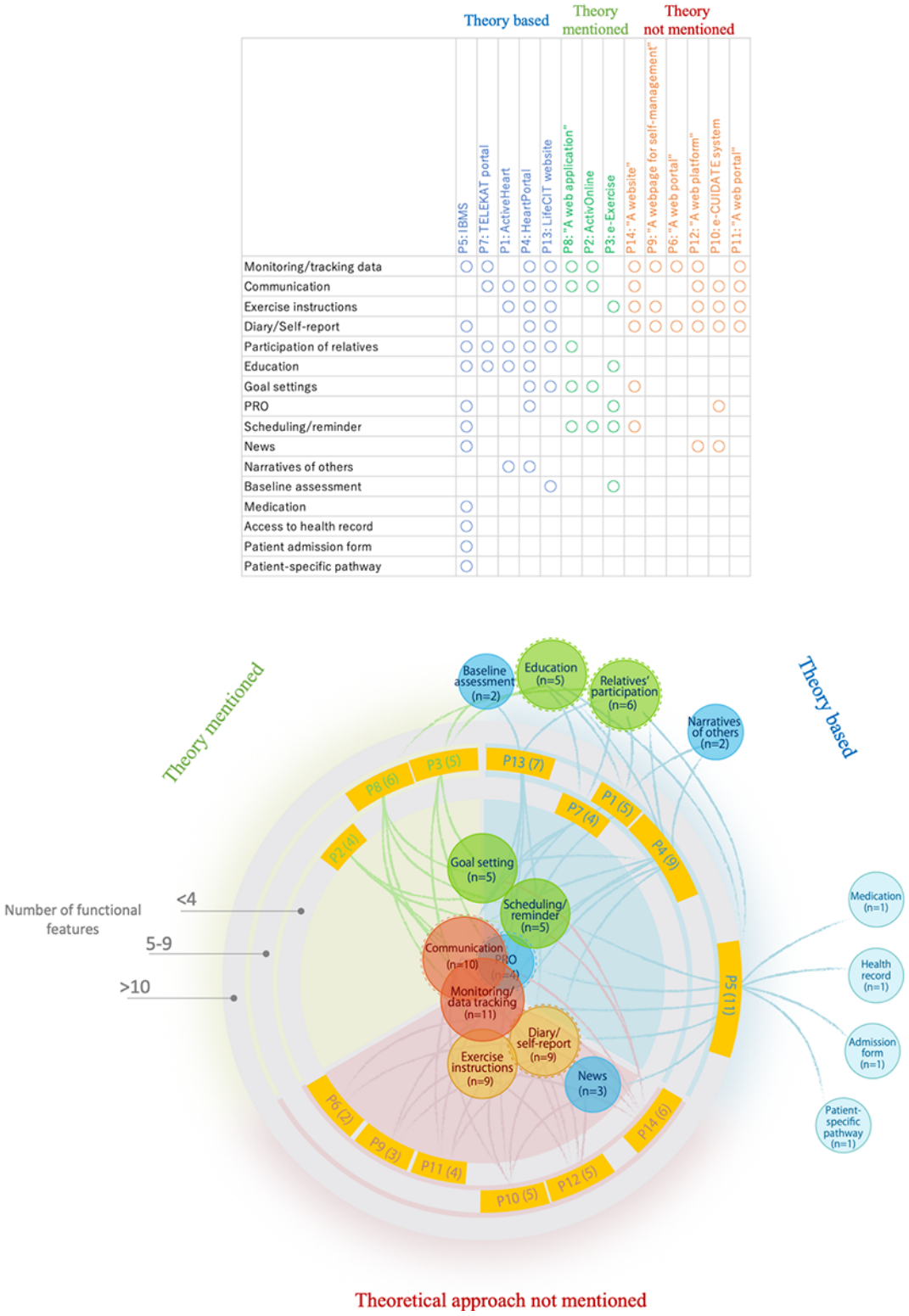


Functional Features of Telerehabilitation Portals

Five portals were for patients with chronic obstructive pulmonary disease (COPD) [30-43], 3 were for patients with cardiovascular diseases [44-49], 2 for multiple sclerosis [50,51], and 1 each for patients with osteoarthritis [52-56], cystic fibrosis [57-59], and stroke [60], and breast cancer survivors [61-64]. These 14 portals were implemented with multiple functions. Monitoring/data tracking functions and communication platforms were the most common features, provided by 11 and 10 portals, respectively. Monitoring/tracking functions foster self-management skills in patients by gaining knowledge and awareness of their health and active involvement in rehabilitation programs [30,37,46]. Communication functions take various forms, including text messaging systems, web forums, and videoconferencing. Exercise instruction function with audiovisual contents was emphasized in 9 portals, and diary/self-report function was found in 9 portals, being used for patients to share their rehabilitation reports and experience with health care team members. Six portals allowed relatives to participate in the rehabilitation program with permission of the patient. Education was also frequently found with 5 portals. Patient-reported outcomes (PROs) data and goal settings can be collected or determined through specifically designed systems [55], as well as through diary or communication functions [48,50,64]. Two portals embedded audiovisual contents about

other patients and their family members talking about their experiences of rehabilitation (narrative of others) [46,48]. Although monitoring is one of the major functions, portals often require assistance from other digital tools [30,35,41,44,50] or extra effort of manual operations [49,57]. The most dominant security measure was an individual, password-protected system. Furthermore, ActivOnline protects user data and privacy by employing security measures (eg, 128-bit SSL security) to provide encrypted data transfer [59], and "web application" (portal 8) employed a 2-factor log-in system with participants' private mobile phone [49]. The summary of functional features and the 14 portals are shown in Table 1 and Figure 2. Detailed descriptions of each portal are given in Multimedia Appendix 1.

Figure 2. A summary of functional features and the 14 portals. A colored circle indicates a functional feature. The number in a circle denotes the number of portals that employ the functional feature. A colored circle with a dotted line is associated with the principle of patient-centered care. Yellow boxes on a gray circle indicate a web portal. The number in parenthesis denotes the number of functional features the portal implemented. P1: ActiveHeart; P2: ActivOnline; P3: e-Exercise; P4: HeartPortal; P5: IBMS; P6: "A web portal"; P7: TELEKAT portal; P8: "A web application", P9: "A webpage for self-management"; P10: e-CUIDATE system; P11: "A web portal"; P12: "A web platform; P13: LifeCIT website; P14: "A website". IBMS: Integrated Care Portal Multiple Sclerosis; TELEKAT: telehomecare, chronic patients, and the integrated health care system; P: portal; PROs: patient reported outcomes.



Some functional features are considered to be associated with the principle of patient-centered care. Respect for patients' values can be attempted by PRO, diary function, and the participatory design process, which several portals followed [36,45,47,50]. Communication and education functions can provide coordination and integration of care, and the function of the relatives' participation is associated with involvement of family and friends.

Table 1. A list of included studies.

Portal name and references	Project name	Diseases	Functional features	Required additional systems
Portal 1: ActiveHeart	TTP[a]	Heart disease	Communication, exercise instructions, education, participation of relatives, narratives of others[b]	The shared care platform, MyMedic, Fitbit, sphygmomanometer, scale
Dinesen et al [44]				
Dinesen et al [45]				
Melholt et al [46]				
Portal 2: ActivOnline	ActionPACT[c]	Cystic fibrosis (later used for COPD[d] and bronchiectasis)	Monitoring/tracking data, communication, goal settings, scheduling/reminder	Step counter (Fitbit, mobile phone or pedometer)
Cox et al [57]				
Liacos et al [58]				
Cox et al [59]				
Portal 3: e-Exercise	e-Exercise	Knee/hip osteoarthritis	Exercise instructions, education, PROs[e], scheduling/reminder, baseline assessment	Face-to-face sessions
Kloek et al [52]				
Kloek et al [53]				
Kloek et al [54]				
de Vries et al [55]				
Bossen et al [56]				
Portal 4: HeartPortal	Future patient	Heart failure	Monitoring/tracking data, communication, exercise instructions, diary, education, participation of relatives, goal settings (via communication and diary function), PROs, narratives of others	Sphygmomanometer, scale, data transmitter, step counters, sleep sensor, iPad
Joensson et al [47]				
Dinesen et al [48]				
Portal 5: IBMS[f]	—[g]	Multiple sclerosis	Monitoring/tracking data, diary, education, participation of relatives, PROs (via diary function), scheduling/reminder, news, medication, access to a health care record (multiple sclerosis care record), patient administration form, patient-specific pathways	MSDS[3Dh], multiple sclerosis case record
Voigt et al [50]				
Portal 6: "A web portal"	—	COPD	Monitoring/tracking data, diary	Activity coach (smartphone and accelerometer)
Tabak et al [30]				
Portal 7: TELEKAT[i] portal	TELEKAT	COPD	Monitoring/tracking data, communication, education (via communication function), participation of relatives	MyMedic/MyMedicPlus, scale, sphygmomanometer, oximeter, spirometer
Dinesen et al [31]				
Haesum et al [32]				
Dinesen et al [33]				
Jensen et al [34]				
Dinesen et al [35]				
Dinesen et al [36]				

Portal name and references	Project name	Diseases	Functional features	Required additional systems
Huniche et al [37]				
Portal 8: "A web application"	Smart-CareCAD	Coronary artery disease	Monitoring/tracking data, communication, participation of relatives, goal setting, scheduling/reminder	Mio alpha, hip-worn accelerometer
Brouwers et al [49]				
Portal 9: "A webpage for self-management"	—	COPD	Monitoring/tracking data, exercise instructions, diary	Treadmill, iPad, iPad holder, web conference application, pulse oximeter
Zanaboni et al [38]				
Hoaas et al [39]				
Zanaboni et al [40]				
Portal 10: e-CUIDATE system	e-CUIDATE	Breast cancer survivors	Communication, exercise instructions, diary, PROs (via communication function), news	Telephone call
Ariza-Garcia et al [61]				
Galiano-Castillo et al [62]				
Galiano-Castillo et al [63]				
Galiano-Castillo et al [64]				
Portal 11: "A web portal"	—	COPD	Monitoring/tracking data, communication, exercise instructions, diary	Activity coach (smartphone and accelerometer)
Tabak et al [41]				
Portal 12": "A web platform"	—	Multiple sclerosis	Monitoring/tracking data, communication, exercise instructions, diary, news	An ad hoc tracking system for determination of posture, ToF camera[j]
Eguiluz-Perez and Garcia-Zapirain [51]				
Portal 13: LifeCIT[k] website	LifeCIT	Poststroke	Monitoring/tracking data, communication, exercise instructions, diary, participation of relatives, goal setting, baseline assessment	C-Mitt (a glove to restrict functional hand movement)
Burridge et al [60]				
Portal 14: "A website"	iTrain	COPD	Monitoring/tracking data, communication, exercise instructions, diary, goal setting, scheduling/reminder	Treadmill, iPad, iPad holder, web conference application, pulse oximeter
Hoaas et al [42]				
Zanaboni et al [43]				

[a]TTP: teledialog telerehabilitation program.

[b]Interviews, stories, and experiences of other patients or relatives who have the same disease.

[c]ActionPACT: the active online physical activity in the cystic fibrosis trial.

[d]COPD: chronic obstructive pulmonary disease.

[e]PROs: patient-reported outcomes.

[f]IBMS: integrated care portal multiple sclerosis.

[g]Not specified.

[h]Integration of the multiple sclerosis documentation system.

[i]TELEKAT: telehomecare, chronic patients, and the integrated health care system.

[j]Time-of-flight camera.

[k]Constraint-induced therapy.

Theoretical Approaches of Telerehabilitation Portals

Theoretical approaches were observed in 8 portals. Among them, *eHealth literacy* described by Norman and Skinner [65], Wenger's *Communities of Practice* [66], and *Self-Determination Theory* [67] were used for both development process and qualitative usability evaluation [31,44,47,48]. The model of *eHealth literacy* was first used in the qualitative data analysis in ActiveHeart [46], and then employed in HeartPortal to foster participants' empowerment [48]. *Communities of practice* focus on people who share a common interest or concern and interact regularly, with learning taking place during interpersonal interactions [66]. The theory influenced the design of portals as facilitating communication and knowledge-sharing functions among participants and patients with chronic diseases in the ActiveHeart and telehomecare, and the integrated health care system (TELEKAT) portals [31,44]. The Self-Determination Theory takes into account intrinsic and extrinsic motivation with 3 basic human needs: autonomy, competence, and relatedness [67]. HeartPortal includes a graphical function that displays charts of each participant's physical data [48]. Participants who monitor their condition using detailed charts are expected to gain autonomy and competence. Relatives' participation upon patient's permission is included in several portals, and it is in the context of relatedness in HeartPortal. *Pathway-based care model*–implemented integrated care portal multiple sclerosis (IBMS) is an approach that provides a patient with a complete picture of his/her disease progression and the current state of an evidence-based treatment strategy [50,68]. By using this model, milestones of the treatment are defined, and shared decision making between patients and the multidisciplinary care team is expected to be induced [50]. *Person-based approach* for designing successful digital interventions [69] was the method used to guide the development phase of LifeCIT's website [60].

Some studies were not "theory based" but introduced theories in their study concept, thus we considered them as "theory mentioned." *Cognitive behavior principles* were mentioned in 3 studies [49,53,57], where 1 study [49] briefly mentions theory linking to relapse prevention. *Operant conditioning* was associated with time-contingent exercise activity in e-Exercise [56]. On the contrary, "A webpage for self-management" (portal 9) stated their study was empirically based. Moreover, e-Exercise, "A web portal" (portal 11), and "A website" (portal 14) used empirical data of their previous trials [41,56].

The BCTs and MTs were identified by the authors' descriptions in the included articles, as well as referring to a taxonomy of BCTs proposed by Abraham and Michie [70]. The incorporation of several BCTs was reported to be more effective than interventions that incorporated fewer techniques [71]. Relatives' participation and goal settings were commonly found in portals that took theoretical approaches. The general purpose of BCTs is to better support patients shifting to a healthier lifestyle. Note that a technique mentioned as BCT in one study can be implemented in another study without being stated as a BCT purpose.

Mode of delivery is also known to influence behavior [71]. It is obvious that all 14 portals are internet-based programs, of which ActivOnline, "A web application" (portal 8), e-CUIDATE, and "A website" (portal 14) chose this mode of delivery with the expectation of better treatment results compared with conventional center-based treatment [43,49,57,61]. As a synchronous communication, 2 portals embedded a videoconferencing function, which is used for consultations rather than exercise instructions. Asynchronous physical activity was intentionally chosen for better adherence to the exercise in several studies [39,41,43,56,64]. Although 2 interventions applied synchronous group exercises in addition to asynchronous exercises thorough a videoconferencing application apart from a portal [39,43], no functional features of the 14 portals were used for synchronous exercise instructions. The LifeCIT website has gaming contents for poststroke exercise [60]. While the portals that took the theoretical approach tend to have more BCTs, the number of modes of delivery did not differ between portals with or without a theoretical approach. Figure 3 summarizes the theoretical approaches, BCTs, MTs, and modes of delivery employed in the 14 portals.

Figure 3. Theoretical approaches, BCTs, MTs, and modes of delivery employed in the 14 portals. Yellow square, theory applied; light yellow squares, theory or model mentioned without specifies on the mode of implication; green squares, BCT applied; light green squares, BCT mentioned without specifies on the mode of implication; green circles, through an additional system; orange squares, mode of delivery applied; orange circles, through an additional system including another digital platform used simultaneously, video consultation with HCP, or face-to-face session as part of the intervention; I: theory based; II: theory mentioned; III: theory not mentioned; BCT: behavior change technique; HCP: health care provider; IBMS: integrated care portal multiple sclerosis; MT: motivational technique; MoD: mode of delivery; TELEKAT: telehomecare, chronic patients, and the integrated health care system.

Discussion

Principal Findings

The objective of this scoping review was to investigate functional features and theoretical approaches of web portals developed for telerehabilitation programs in patients with chronic diseases, as well as any characteristics that can be observed through the investigation. As a result, monitoring/tracking data were found to be the most common function in telerehabilitation portals. Data monitoring by patients allowed them to gain self-management skills, which is a key for controlling chronic diseases [11]. Moreover, portal 9 was called by the authors "A webpage for self-management." Those results indicate that fostering self-management skills is one of the characteristic features of telerehabilitation web portals. Another characteristic feature can be observed through a communication function, because patients with chronic symptoms often require a multidisciplinary care team, including patients' relatives [48,50]. Facilitating physical activities is another characteristic function of telerehabilitation portals, and exercise programs are often individually modified [41,43,49,56,60,64]. Nine portals embedded either audiovisual exercise instructions or exercise game contents, whereas ActivOnline, "A web portal" (portal 6), and "A web application" (portal 8) promote physical activity by a monitoring function. The TELEKAT portal and IBMS do not promote exercises directly; however, exercise promotion is included in a part of the TELEKAT intervention. As such, the web portals not only present rehabilitation instructions but also facilitate rehabilitation activities through various functional features. In other words, the underlying concept of web portals is to lower the hurdle and promote rehabilitation to become part of patients' everyday lives.

Taken together, we propose several key concepts that can be addressed in the development of a telerehabilitation web portal. The advantage of web portals is their ability to unify and provide various information, which can be shared by many stakeholders, including patients' relatives. By contrast, acquiring data and real-time correspondence often demand an extra system. Regarding the weakness of data acquisition and real-time correspondence, providing patients with a tablet or smartphone with a Bluetooth connection may help solve such problems, as exemplified in some studies [30,40,43,48]. Asynchronous than synchronous rehabilitation is more preferrable for web portals, as it is regarded as less time restrictive and more able to access a rehabilitation program [72-75]. This may suit rehabilitation participants who struggle with disruption to their established routine. The principle of patient-centered care may guide development of the digital platform. Consequently, conducting a participatory design process may be desirable. Theoretical approaches are recommended because theories and models can be used not only during the developing phase but also to analyze outcomes. Moreover, using a theoretical approach enables an intervention to be mapped to existing knowledge. Functional features may be developed upon the association of BCTs or MTs. In addition, other BCTs that were not observed in the 14 portals may be included in future telerehabilitation portals, for example, prompt barrier identification or mindfulness techniques [70,76]. Many different kinds of digital platforms are currently available for telerehabilitation. Reasons for choosing certain platforms, such as web applications, iOS/Android apps, and eHealth device–associated applications, are considered to be knowledge gaps, which should be studied in the future. Differences in platforms can be analyzed from the perspective of data privacy issues. Because web portals often contain highly personal information, choosing a web application involves an important security strategy. Security updates of web applications are done by system providers, and, unlike native apps, little action is required from the user side. This is a great advantage for participants, especially those unfamiliar with digital technologies.

Theoretical approaches were found in more than half of the portals analyzed, and BCTs were often mentioned in those studies. Current knowledge of BCTs, however, may not be sufficient to understand BCTs used for telerehabilitation platforms. For instance, although monitoring/tracking data are associated with BCTs and MTs in several web portal articles [30,31,59], the function is not monitoring behavior itself but rather monitoring physical condition or exercise performance. Developing a new analytical method for theories, BCTs, MTs, and mode of delivery used in telerehabilitation may provide evidence on which to base selection of particular combinations.

Web portals are not only practical tools to collect data or supply exercise instructions. If remote monitoring by health care professionals is the only goal, portals may not be necessary to be developed. Patients with chronic diseases often require lifestyle changes. Through patient-centered functions, web portals deliver various useful support for patients with chronic diseases. Web portals also assist rehabilitation to become part of patients' daily routines, which is considered a key to guiding successful disease management [44]. Web portals have the potential to facilitate patients in applying a new, healthy lifestyle in accordance with their symptoms and help participants master their own diseases in everyday life.

Limitations

International consensus about web portals for telerehabilitation is scarce. To effectively extract the characteristics, rather strict inclusion criteria were applied in this study. As mentioned in the "Methods" section, portals developed for purposes other than telerehabilitation were excluded, as were clinician portals, portals targeting other than chronic diseases, web applications designed for a single purpose, and web services with limited time of access. In this manner, we were able to bring out some characteristic features of web portals for telerehabilitation. We believe that our findings can be used in comparative studies of digital platforms of telerehabilitation in future investigations. Another limitation is that some of the projects described in this review are currently ongoing, suggesting that the functions of their portals may be improved in the future. Finally, this study did not include the cost of utilities, or patients' and health care professionals' experiences and perspectives when using the portals.

Conclusions

Telerehabilitation has several advantages, including reducing the need to visit clinics by patients vulnerable to infectious diseases. The global COVID-19 pandemic has increased the need for telerehabilitation. Web portals have the potential to be a core digital component for patient-centered telerehabilitation. The common functional features of 14 web portals studied in this scoping review were monitoring/tracking data provided by 11 portals; communication, 10; exercise instructions, 9; diary/self-report functions, 9; relatives' participation, 6; goal settings, 5; and education, 5. Although different functional features addressed various purposes, the underlying concept was to facilitate rehabilitation to become a part of patients' everyday lives. Web portals were able to unify and display multiple types of data and could effectively provide various types of information. Asynchronous correspondence was more favorable than synchronous, real-time interactions. Data acquisition often demanded other digital tools. As much as 8 out of 14 portals employed theoretical approaches to some degree, and various patient-centered functions were observed. As is usual in web applications, security updates were the responsibility of service providers, and thus such platforms may be especially suitable for participants who are unfamiliar with digital technologies. These findings suggested that web portals for telerehabilitation not only provide entrance into rehabilitation programs but also reinforce patient-centered treatment, adherence to rehabilitation, and lifestyle changes over time.

Acknowledgments

This work was supported by a JSPS KAKENHI grant (20H04055). We thank Kenta Takahashi of Juntendo University Information Center for sharing his knowledge of computer science. We also thank Ai Nakahata of Juntendo University Academic Media Center (Library) for her professional assistance. Finally, we thank David Price and Swetha Anand for their help with the editorial preparation of the manuscript.

Conflicts of Interest

TT, NK, TK, and HD are affiliated with a department funded by Philips Japan; Asahi KASEI Corporation; Inter Reha Co, Ltd; and Toho Holdings Co, Ltd based on collaborative research agreements.

Multimedia Appendix 1
Descriptions of each portal.
[DOCX File , 34 KB - jmir_v24i1e27759_app1.docx]

References

1. World Health Organization (WHO). The World Health Report 2002 - Reducing Risks, Promoting Healthy Life. Geneva, Switzerland: WHO; 2002.
2. World Health Organization (WHO). Preventing Chronic Diseases - A Vital Investment. Geneva, Switzerland: WHO; 2005.
3. Richardson CR, Franklin B, Moy ML, Jackson EA. Advances in rehabilitation for chronic diseases: improving health outcomes and function. BMJ 2019 Jun 17;365:l2191. [doi: 10.1136/bmj.l2191] [Medline: 31208954]
4. Keating A, Lee A, Holland AE. What prevents people with chronic obstructive pulmonary disease from attending pulmonary rehabilitation? A systematic review. Chron Respir Dis 2011 May;8(2):89-99 [FREE Full text] [doi: 10.1177/1479972310393756] [Medline: 21596892]
5. Moulson N, Bewick D, Selway T, Harris J, Suskin N, Oh P, et al. Cardiac Rehabilitation During the COVID-19 Era: Guidance on Implementing Virtual Care. Can J Cardiol 2020;19:1317-1321. [doi: 10.1016/j.cjca.2020.06.006] [Medline: 32553606]
6. Leocani L, Diserens K, Moccia M, Caltagirone C. Disability through COVID‐19 pandemic: neurorehabilitation cannot wait. Eur J Neurol 2020 Jun 25:e50-e51. [doi: 10.1111/ene.14320] [Medline: 32402100]
7. Czeisler M, Marynak K, Clarke KE, Salah Z, Shakya I, Thierry JM, et al. Delay or Avoidance of Medical Care Because of COVID-19-Related Concerns - United States, June 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 11;69(36):1250-1257 [FREE Full text] [doi: 10.15585/mmwr.mm6936a4] [Medline: 32915166]
8. Brennan D, Tindall L, Theodoros D, Brown J, Campbell M, Christiana D, et al. A blueprint for telerehabilitation guidelines. Int J Telerehabil 2010;2(2):31-34 [FREE Full text] [doi: 10.5195/ijt.2010.6063] [Medline: 25945175]
9. Cox N, Dal Corso S, Hansen H, McDonald C, Hill C, Zanaboni P, et al. Telerehabilitation for chronic respiratory disease. Cochrane Database Syst Rev 2021 Jan 29;1:CD013040. [doi: 10.1002/14651858.CD013040.pub2] [Medline: 33511633]
10. Laver K, Adey-Wakeling Z, Crotty M, Lannin N, George S. Telerehabilitation services for stroke. Cochrane Database Syst Rev 2020;1(1):CD010255. [doi: 10.1002/14651858.cd010255.pub3] [Medline: 32002991]
11. Brouwers RWM, van Exel HJ, van Hal JMC, Jorstad HT, de Kluiver EP, Kraaijenhagen RA, Committee for Cardiovascular PreventionCardiac Rehabilitation of the Netherlands Society of Cardiology. Cardiac telerehabilitation as an alternative to centre-based cardiac rehabilitation. Neth Heart J 2020 Sep;28(9):443-451 [FREE Full text] [doi: 10.1007/s12471-020-01432-y] [Medline: 32495296]

12. Choukou M, Maddahi A, Polyvyana A, Monnin C. Digital health technology for Indigenous older adults: A scoping review. Int J Med Inform 2021 Apr;148:104408. [doi: 10.1016/j.ijmedinf.2021.104408] [Medline: 33609927]
13. Saner H, van der Velde E. eHealth in cardiovascular medicine: A clinical update. Eur J Prev Cardiol 2016 Oct;23(2 suppl):5-12. [doi: 10.1177/2047487316670256] [Medline: 27892420]
14. Zasadzka E, Trzmiel T, Pieczyńska A, Hojan K. Modern Technologies in the Rehabilitation of Patients with Multiple Sclerosis and Their Potential Application in Times of COVID-19. Medicina (Kaunas) 2021 May 30;57(6):549 [FREE Full text] [doi: 10.3390/medicina57060549] [Medline: 34070705]
15. Oxford Learner's Dictionaries. URL: https://www.oxfordlearnersdictionaries.com/ [accessed 2022-01-18]
16. Hoogenbosch B, Postma J, de Man-van Ginkel JM, Tiemessen NA, van Delden JJ, van Os-Medendorp H. Use and the Users of a Patient Portal: Cross-Sectional Study. J Med Internet Res 2018 Sep 17;20(9):e262 [FREE Full text] [doi: 10.2196/jmir.9418] [Medline: 30224334]
17. Medical Dictionary. URL: https://medical-dictionary.thefreedictionary.com/ [accessed 2022-01-18]
18. What Is a Patient Portal?. 2021. URL: https://www.healthit.gov/faq/what-patient-portal [accessed 2022-01-18]
19. Van den Bulck SA, Hermens R, Slegers K, Vandenberghe B, Goderis G, Vankrunkelsven P. Designing a Patient Portal for Patient-Centered Care: Cross-Sectional Survey. J Med Internet Res 2018 Oct 01;20(10):e269 [FREE Full text] [doi: 10.2196/jmir.9497] [Medline: 30287416]
20. National Learning Consortium. How to Optimize Patient Portals for Patient Engagement and Meet Meaningful Use Requirements. 2013. URL: https://www.healthit.gov/sites/default/files/nlc_how_to_optimizepatientportals_for_patientengagement.pdf [accessed 2022-01-18]
21. Institute of Medicine (US) Committee on Quality of Health Care in America. Crossing the Quality Chasm: A New Health System for the 21st Century. Washington (DC): National Academies Press; 2001.
22. Gerteis M, Edgman-Levitan S, Daley J, Delbanco TL. Through the Patient's Eyes: Understanding and Promoting Patient-Centered Care. San Francisco, CA: Jossey-Bass; 2002.
23. Conroy SS, Harcum S, Keldsen L, Bever CT. Novel use of existing technology: A preliminary study of patient portal use for telerehabilitation. J Telemed Telecare 2020 Sep 01:1357633X20950172. [doi: 10.1177/1357633X20950172] [Medline: 32869689]
24. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Oct 02;169(7):467-473 [FREE Full text] [doi: 10.7326/M18-0850] [Medline: 30178033]
25. Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist. URL: http://www.prisma-statement.org/documents/PRISMA-ScR-Fillable-Checklist_11Sept2019.pdf [accessed 2022-01-18]
26. Final Version of the Study Protocol. 2021 Jul. URL: https://osf.io/a457t/ [accessed 2022-01-18]
27. PubMed. URL: https://pubmed.ncbi.nlm.nih.gov/ [accessed 2022-01-18]
28. Web of Science. URL: https://apps.webofknowledge.com/ [accessed 2022-01-18]
29. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
30. Tabak M, Vollenbroek-Hutten MM, van der Valk PD, van der Palen J, Hermens HJ. A telerehabilitation intervention for patients with Chronic Obstructive Pulmonary Disease: a randomized controlled pilot trial. Clin Rehabil 2014 Jun 29;28(6):582-591. [doi: 10.1177/0269215513512495] [Medline: 24293120]
31. Dinesen B, Huniche L, Toft E. Attitudes of COPD patients towards tele-rehabilitation: a cross-sector case study. Int J Environ Res Public Health 2013 Nov 15;10(11):6184-6198 [FREE Full text] [doi: 10.3390/ijerph10116184] [Medline: 24247995]
32. Haesum L, Soerensen N, Dinesen B, Nielsen C, Grann O, Hejlesen O, et al. Cost-utility analysis of a telerehabilitation program: a case study of COPD patients. Telemed J E Health 2012 Nov;18(9):688-692. [doi: 10.1089/tmj.2011.0250] [Medline: 23020647]
33. Dinesen B, Haesum LK, Soerensen N, Nielsen C, Grann O, Hejlesen O, et al. Using preventive home monitoring to reduce hospital admission rates and reduce costs: a case study of telehealth among chronic obstructive pulmonary disease patients. J Telemed Telecare 2012 Jun 31;18(4):221-225. [doi: 10.1258/jtt.2012.110704] [Medline: 22653618]
34. Jensen MH, Cichosz SL, Dinesen B, Hejlesen OK. Moving prediction of exacerbation in chronic obstructive pulmonary disease for patients in telecare. J Telemed Telecare 2012 Mar 20;18(2):99-103. [doi: 10.1258/jtt.2011.110607] [Medline: 22267305]
35. Dinesen B, Andersen S, Hejlesen O, Toft E. Interaction between COPD patients and healthcare professionals in a cross-sector tele-rehabilitation programme. Stud Health Technol Inform 2011;169:28-32. [Medline: 21893708]
36. Dinesen B, Seeman J, Gustafsson J. Development of a program for tele-rehabilitation of COPD patients across sectors: co-innovation in a network. Int J Integr Care 2011 Jan;11:e012 [FREE Full text] [doi: 10.5334/ijic.582] [Medline: 21637709]
37. Huniche L, Dinesen B, Grann O, Toft E, Nielsen C. Empowering patients with COPD using Tele-homecare technology. Stud Health Technol Inform 2010;155:48-54. [Medline: 20543309]

38. Zanaboni P, Hoaas H, Aarøen Lien L, Hjalmarsen A, Wootton R. Long-term exercise maintenance in COPD via telerehabilitation: a two-year pilot study. J Telemed Telecare 2016 Jul 09;23(1):74-82. [doi: 10.1177/1357633x15625545] [Medline: 26888420]
39. Hoaas H, Andreassen HK, Lien LA, Hjalmarsen A, Zanaboni P. Adherence and factors affecting satisfaction in long-term telerehabilitation for patients with chronic obstructive pulmonary disease: a mixed methods study. BMC Med Inform Decis Mak 2016 Feb 25;16(1):26 [FREE Full text] [doi: 10.1186/s12911-016-0264-9] [Medline: 26911326]
40. Zanaboni P, Lien LA, Hjalmarsen A, Wootton R. Long-term telerehabilitation of COPD patients in their homes: interim results from a pilot study in Northern Norway. J Telemed Telecare 2013 Oct;19(7):425-429. [doi: 10.1177/1357633X13506514] [Medline: 24218358]
41. Tabak M, Brusse-Keizer M, van der Valk P, Hermens H, Vollenbroek-Hutten M. A telehealth program for self-management of COPD exacerbations and promotion of an active lifestyle: a pilot randomized controlled trial. Int J Chron Obstruct Pulmon Dis 2014;9:935-944 [FREE Full text] [doi: 10.2147/COPD.S60179] [Medline: 25246781]
42. Hoaas H, Zanaboni P, Hjalmarsen A, Morseth B, Dinesen B, Burge A, et al. Seasonal variations in objectively assessed physical activity among people with COPD in two Nordic countries and Australia: a cross-sectional study. Int J Chron Obstruct Pulmon Dis 2019;14:1219-1228 [FREE Full text] [doi: 10.2147/COPD.S194622] [Medline: 31239657]
43. Zanaboni P, Dinesen B, Hjalmarsen A, Hoaas H, Holland AE, Oliveira CC, et al. Long-term integrated telerehabilitation of COPD Patients: a multicentre randomised controlled trial (iTrain). BMC Pulm Med 2016 Aug 22;16(1):126 [FREE Full text] [doi: 10.1186/s12890-016-0288-z] [Medline: 27549782]
44. Dinesen B, Nielsen G, Andreasen JJ, Spindler H. Integration of Rehabilitation Activities Into Everyday Life Through Telerehabilitation: Qualitative Study of Cardiac Patients and Their Partners. J Med Internet Res 2019 Apr 15;21(4):e13281 [FREE Full text] [doi: 10.2196/13281] [Medline: 30985284]
45. Dinesen B, Spindler H. The Use of Telerehabilitation Technologies for Cardiac Patients to Improve Rehabilitation Activities and Unify Organizations: Qualitative Study. JMIR Rehabil Assist Technol 2018 Nov 19;5(2):e10758 [FREE Full text] [doi: 10.2196/10758] [Medline: 30455168]
46. Melholt C, Joensson K, Spindler H, Hansen J, Andreasen JJ, Nielsen G, et al. Cardiac patients' experiences with a telerehabilitation web portal: Implications for eHealth literacy. Patient Educ Couns 2018 May;101(5):854-861. [doi: 10.1016/j.pec.2017.12.017] [Medline: 29305064]
47. Joensson K, Melholt C, Hansen J, Leth S, Spindler H, Olsen M, et al. Listening to the patients: using participatory design in the development of a cardiac telerehabilitation web portal. Mhealth 2019;5:33 [FREE Full text] [doi: 10.21037/mhealth.2019.08.06] [Medline: 31620460]
48. Dinesen B, Dittmann L, Gade J, Jørgensen CK, Hollingdal M, Leth S, et al. "Future Patient" Telerehabilitation for Patients With Heart Failure: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2019 Sep 19;8(9):e14517 [FREE Full text] [doi: 10.2196/14517] [Medline: 31538944]
49. Brouwers RWM, Kraal JJ, Traa SCJ, Spee RF, Oostveen LMLC, Kemps HMC. Effects of cardiac telerehabilitation in patients with coronary artery disease using a personalised patient-centred web application: protocol for the SmartCare-CAD randomised controlled trial. BMC Cardiovasc Disord 2017 Dec 31;17(1):46 [FREE Full text] [doi: 10.1186/s12872-017-0477-6] [Medline: 28143388]
50. Voigt I, Benedict M, Susky M, Scheplitz T, Frankowitz S, Kern R, et al. A Digital Patient Portal for Patients With Multiple Sclerosis. Front Neurol 2020;11:400 [FREE Full text] [doi: 10.3389/fneur.2020.00400] [Medline: 32670174]
51. Eguiluz-Perez G, Garcia-Zapirain B. Comprehensive verticality analysis and web-based rehabilitation system for people with multiple sclerosis with supervised medical monitoring. Biomed Mater Eng 2014;24(6):3493-3502. [doi: 10.3233/BME-141175] [Medline: 25227062]
52. Kloek CJ, Bossen D, de Vries HJ, de Bakker DH, Veenhof C, Dekker J. Physiotherapists' experiences with a blended osteoarthritis intervention: a mixed methods study. Physiother Theory Pract 2020 May 28;36(5):572-579. [doi: 10.1080/09593985.2018.1489926] [Medline: 29952687]
53. Kloek CJJ, van Dongen JM, de Bakker DH, Bossen D, Dekker J, Veenhof C. Cost-effectiveness of a blended physiotherapy intervention compared to usual physiotherapy in patients with hip and/or knee osteoarthritis: a cluster randomized controlled trial. BMC Public Health 2018 Aug 31;18(1):1082 [FREE Full text] [doi: 10.1186/s12889-018-5975-7] [Medline: 30170586]
54. Kloek CJJ, Bossen D, Spreeuwenberg PM, Dekker J, de Bakker DH, Veenhof C. Effectiveness of a Blended Physical Therapist Intervention in People With Hip Osteoarthritis, Knee Osteoarthritis, or Both: A Cluster-Randomized Controlled Trial. Phys Ther 2018 Jul 01;98(7):560-570 [FREE Full text] [doi: 10.1093/ptj/pzy045] [Medline: 29788253]
55. de VHJ, Kloek CJJ, de BDH, Dekker J, Bossen D, Veenhof C. Determinants of Adherence to the Online Component of a Blended Intervention for Patients with Hip and/or Knee Osteoarthritis: A Mixed Methods Study Embedded in the e-Exercise Trial. Telemed J E Health 2017 Dec;23(12):1002-1010. [doi: 10.1089/tmj.2016.0264] [Medline: 28525310]
56. Bossen D, Kloek C, Snippe HW, Dekker J, de Bakker D, Veenhof C. A Blended Intervention for Patients With Knee and Hip Osteoarthritis in the Physical Therapy Practice: Development and a Pilot Study. JMIR Res Protoc 2016 Feb 24;5(1):e32 [FREE Full text] [doi: 10.2196/resprot.5049] [Medline: 26912378]
57. Cox NS, Eldridge B, Rawlings S, Dreger J, Corda J, Hauser J, Youth Activity Unlimited – A Strategic Research Centre of the UK Cystic Fibrosis Trust. A web-based intervention to promote physical activity in adolescents and young adults with

cystic fibrosis: protocol for a randomized controlled trial. BMC Pulm Med 2019 Dec 19;19(1):253 [FREE Full text] [doi: 10.1186/s12890-019-0942-3] [Medline: 31856791]
58. Liacos A, Burge A, Cox N, Holland AE. Promoting Physical Activity Using the Internet: Is It Feasible and Acceptable for Patients With Chronic Obstructive Pulmonary Disease and Bronchiectasis? J Aging Phys Act 2018 Jul 01;26(3):372-381. [doi: 10.1123/japa.2017-0123] [Medline: 28952865]
59. Cox NS, Alison JA, Button BM, Wilson JW, Holland AE. Feasibility and acceptability of an internet-based program to promote physical activity in adults with cystic fibrosis. Respir Care 2015 Mar;60(3):422-429 [FREE Full text] [doi: 10.4187/respcare.03165] [Medline: 25425703]
60. Burridge JH, Lee ACW, Turk R, Stokes M, Whitall J, Vaidyanathan R, et al. Telehealth, Wearable Sensors, and the Internet: Will They Improve Stroke Outcomes Through Increased Intensity of Therapy, Motivation, and Adherence to Rehabilitation Programs? J Neurol Phys Ther 2017 Jul;41 Suppl 3:S32-S38. [doi: 10.1097/NPT.0000000000000183] [Medline: 28628594]
61. Ariza-Garcia A, Lozano-Lozano M, Galiano-Castillo N, Postigo-Martin P, Arroyo-Morales M, Cantarero-Villanueva I. A Web-Based Exercise System (e-CuidateChemo) to Counter the Side Effects of Chemotherapy in Patients With Breast Cancer: Randomized Controlled Trial. J Med Internet Res 2019 Jul 24;21(7):e14418 [FREE Full text] [doi: 10.2196/14418] [Medline: 31342907]
62. Galiano-Castillo N, Arroyo-Morales M, Lozano-Lozano M, Fernández-Lao C, Martín-Martín L, Del-Moral-Ávila R, et al. Effect of an Internet-based telehealth system on functional capacity and cognition in breast cancer survivors: a secondary analysis of a randomized controlled trial. Support Care Cancer 2017 Nov 22;25(11):3551-3559. [doi: 10.1007/s00520-017-3782-9] [Medline: 28639097]
63. Galiano-Castillo N, Cantarero-Villanueva I, Fernández-Lao C, Ariza-García A, Díaz-Rodríguez L, Del-Moral-Ávila R, et al. Telehealth system: A randomized controlled trial evaluating the impact of an internet-based exercise intervention on quality of life, pain, muscle strength, and fatigue in breast cancer survivors. Cancer 2016 Oct 15;122(20):3166-3174 [FREE Full text] [doi: 10.1002/cncr.30172] [Medline: 27332968]
64. Galiano-Castillo N, Ariza-García A, Cantarero-Villanueva I, Fernández-Lao C, Díaz-Rodríguez L, Legerén-Alvarez M, et al. Telehealth system (e-CUIDATE) to improve quality of life in breast cancer survivors: rationale and study protocol for a randomized clinical trial. Trials 2013 Jun 22;14:187 [FREE Full text] [doi: 10.1186/1745-6215-14-187] [Medline: 23799886]
65. Norman CD, Skinner HA. eHealth Literacy: Essential Skills for Consumer Health in a Networked World. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]
66. Wenger E. Communities of Practice: Learning, Meaning, and Identity (Learning in Doing: Social, Cognitive and Computational Perspectives). Cambridge, MA: Cambridge University Press; 1999.
67. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American Psychologist 2000;55(1):68-78. [doi: 10.1037//0003-066X.55.1.68]
68. Schlieter H, Benedict M, Gand K, Burwitz M. Towards Adaptive Pathways: Reference Architecture for Personalized Dynamic Pathways. 2017 Presented at: IEEE 19th Conference on Business Informatics (CBI); July 24-27, 2017; Thessaloniki, Greece p. 359-368. [doi: 10.1109/cbi.2017.55]
69. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
70. Abraham C, Michie S. A taxonomy of behavior change techniques used in interventions. Health Psychol 2008 May;27(3):379-387. [doi: 10.1037/0278-6133.27.3.379] [Medline: 18624603]
71. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
72. Bini S, Mahajan J. Clinical outcomes of remote asynchronous telerehabilitation are equivalent to traditional therapy following total knee arthroplasty: A randomized control study. J Telemed Telecare 2017 Feb;23(2):239-247. [doi: 10.1177/1357633X16634518] [Medline: 26940798]
73. Naeemabadi M, Dinesen B, Andersen OK, Madsen NK, Simonsen OH, Hansen J. Developing a telerehabilitation programme for postoperative recovery from knee surgery: specifications and requirements. BMJ Health Care Inform 2019 Apr;26(1):e000022 [FREE Full text] [doi: 10.1136/bmjhci-2019-000022] [Medline: 31039126]
74. Naeemabadi M, Søndergaard JH, Klastrup A, Schlünsen AP, Lauritsen REK, Hansen J, et al. Development of an individualized asynchronous sensor-based telerehabilitation program for patients undergoing total knee replacement: Participatory design. Health Informatics J 2020 Dec;26(4):2492-2511 [FREE Full text] [doi: 10.1177/1460458220909779] [Medline: 32175788]
75. Naeemabadi M, Fazlali H, Najafi S, Dinesen B, Hansen J. Telerehabilitation for Patients With Knee Osteoarthritis: A Focused Review of Technologies and Teleservices. JMIR Biomed Eng 2020 Jul 21;5(1):e16991. [doi: 10.2196/16991]
76. Schuman-Olivier Z, Trombka M, Lovas DA, Brewer JA, Vago DR, Gawande R, et al. Mindfulness and Behavior Change. Harv Rev Psychiatry 2020;28(6):371-394 [FREE Full text] [doi: 10.1097/HRP.0000000000000277] [Medline: 33156156]

Abbreviations

ActionPACT: the active online physical activity in the cystic fibrosis trial
BCT: behavior change technique
COPD: chronic obstructive pulmonary disease
IBMS: integrated care portal multiple sclerosis
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-analysis Extension for Scoping Reviews
PROs: patient-reported outcomes
TELEKAT: telehomecare, chronic patients, and the integrated health care system
TTP: teledialog telerehabilitation program

Edited by A Mavragani; submitted 12.02.21; peer-reviewed by P Postigo-Martin, J Kraal, K Kastelic, D Bossen; comments to author 17.04.21; revised version received 08.09.21; accepted 03.12.21; published 27.01.22.

Please cite as:
Morimoto Y, Takahashi T, Sawa R, Saitoh M, Morisawa T, Kagiyama N, Kasai T, Dinesen B, Hollingdal M, Refsgaard J, Daida H Web Portals for Patients With Chronic Diseases: Scoping Review of the Functional Features and Theoretical Frameworks of Telerehabilitation Platforms
J Med Internet Res 2022;24(1):e27759
URL: https://www.jmir.org/2022/1/e27759
doi:10.2196/27759
PMID:35084355

Review

Methods for Human-Centered eHealth Development: Narrative Scoping Review

Hanneke Kip[1,2], PhD; Julia Keizer[1], MSc; Marcia C da Silva[1], MSc; Nienke Beerlage-de Jong[3], PhD; Nadine Köhle[1], PhD; Saskia M Kelders[1,4], PhD

[1]Centre for eHealth and Wellbeing Research, Department of Psychology, Health and Technology, University of Twente, Enschede, Netherlands

[2]Department of Research, Transfore, Deventer, Netherlands

[3]Department of Health Technology & Services Research, University of Twente, Enschede, Netherlands

[4]Optentia Research Focus Area, North-West University, Vanderbijlpark, South Africa

Corresponding Author:
Hanneke Kip, PhD
Centre for eHealth and Wellbeing Research
Department of Psychology, Health and Technology
University of Twente
Drienerlolaan 5
Enschede, 7500 AE
Netherlands
Phone: 31 +31534896536
Email: h.kip@utwente.nl

Abstract

Background: Thorough holistic development of eHealth can contribute to a good fit among the technology, its users, and the context. However, despite the availability of frameworks, not much is known about specific research activities for different aims, phases, and settings. This results in researchers having to reinvent the wheel. Consequently, there is a need to synthesize existing knowledge on research activities for participatory eHealth development processes.

Objective: The 3 main goals of this review are to create an overview of the development strategies used in studies based on the CeHRes (Center for eHealth Research) Roadmap, create an overview of the goals for which these methods can be used, and provide insight into the lessons learned about these methods.

Methods: We included eHealth development studies that were based on the phases and/or principles of the CeHRes Roadmap. This framework was selected because of its focus on participatory, iterative eHealth design in context and to limit the scope of this review. Data were extracted about the type of strategy used, rationale for using the strategy, research questions, and reported information on lessons learned. The most frequently mentioned lessons learned were summarized using a narrative, inductive approach.

Results: In the included 160 papers, a distinction was made between overarching development *methods* (n=10) and *products* (n=7). *Methods* are used to gather new data, whereas *products* can be used to synthesize previously collected data and support the collection of new data. The identified methods were focus groups, interviews, questionnaires, usability tests, literature studies, desk research, log data analyses, card sorting, Delphi studies, and experience sampling. The identified products were prototypes, requirements, stakeholder maps, values, behavior change strategies, personas, and business models. Examples of how these methods and products were applied in the development process and information about lessons learned were provided.

Conclusions: This study shows that there is a plethora of methods and products that can be used at different points in the development process and in different settings. To do justice to the complexity of eHealth development, it seems that multiple strategies should be combined. In addition, we found no evidence for an optimal single step-by-step approach to develop eHealth. Rather, researchers need to select the most suitable research methods for their research objectives, the context in which data are collected, and the characteristics of the participants. This study serves as a first step toward creating a toolkit to support researchers in applying the CeHRes Roadmap to practice. In this way, they can shape the most suitable and efficient eHealth development process.

(J Med Internet Res 2022;24(1):e31858) doi:10.2196/31858

KEYWORDS

eHealth; community-based participatory research; human-centered design; CeHRes Roadmap; internet-based intervention; technological innovations

Introduction

Background

Over the past years, many different types of eHealth technologies have been developed, implemented, and studied in practice. These eHealth technologies, such as web-based interventions or mobile apps, are used to support health, well-being, and health care using technology [1]. Although the *e* in eHealth illustrates the importance of technology, eHealth encompasses much more than merely adding information and communication technology (ICT). It characterizes a novel way of thinking and working, and it changes the way health care is organized [2,3]. eHealth can offer many benefits such as increased access to care, increased efficiency and quality of care, and more ownership and self-management among patients [1]. However, in practice, many of these potential benefits are not achieved. A reason for this is low uptake; many eHealth technologies are not used as often as would be expected [4]. This can be partly explained by a suboptimal fit among the characteristics of a technology, the needs and skills of the users, and the context in which the technology is used [5]. If the content of an eHealth technology does not fit with the structures of an organization and the characteristics of end users, chances of its being successfully used are low [6]. To illustrate, if a web-based intervention requires a lot of reading, it will probably not fit well within an organization that mostly treats patients with low literacy skills. This interrelationship highlights the holistic nature of eHealth, in which technology, people, and context are intertwined. For that purpose, a user-centered, iterative, and multi-method development process in which all stakeholders are actively involved is recommended [7-10]. By means of a thorough development process in which multiple research activities are combined, eHealth that provides added value for its users and context can be realized [11].

Models for eHealth Development

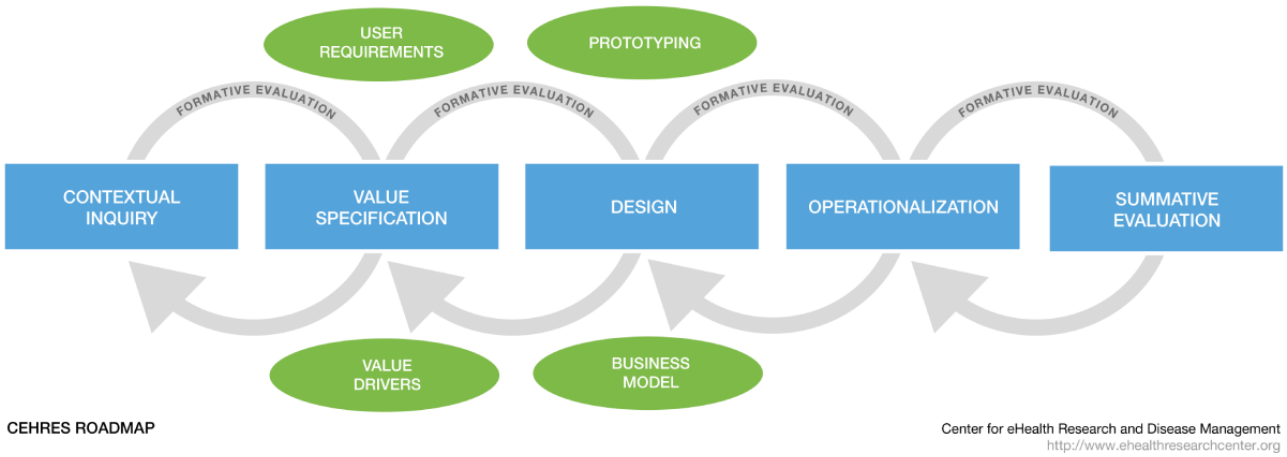
There are multiple frameworks and models that can be used to guide human-centered, iterative development processes of eHealth. Well-known examples are the CeHRes (Center for eHealth Research) Roadmap [2,11]; the person-based approach [7]; the Accelerated Creation-to-Sustainment model [12], Intervention Mapping [9], the Persuasive System Design model [13], and the agile science approach [14]. Although these abstract models offer valuable guidelines and principles, they are not, and should not be, viewed as step-by-step prescriptions of eHealth development [14,15]. Rather, they should be viewed as a framework that researchers and developers use to shape their own development process and select the most appropriate research activities. However, not much is known about which research activities are most suitable for eHealth development within specific types of contexts and participants [16]. Consequently, there might be an availability bias in eHealth development: researchers might mostly use research activities that they are experienced with or those that are often described in literature [17]. However, other less-known research activities might have been a better fit with their research questions and context. To increase knowledge on how to apply development models in practice, existing eHealth frameworks could be supplemented with practical toolkits. Such toolkits could support the operationalization of the more abstract frameworks into specific research activities [15]. They can be based on experiences and lessons learned from earlier research. In this way, they could provide an overview of the kinds of development activities that can be used in the different phases of a specific eHealth development framework and offer guidelines on when and how to use these activities.

Objective

In this study, we aim to create the foundation for a toolkit for a specific eHealth development framework. We provide an overview of research activities for the development of eHealth technology in context. Including studies on all eHealth development projects conducted would be a very time consuming and nearly impossible task. Therefore, this paper focuses only on studies that are based on the CeHRes Roadmap. The CeHRes Roadmap is a much-used structured framework for the development of eHealth technologies [5,16]. The results of this review can support researchers to apply the CeHRes Roadmap in practice by supporting them to select not the most obvious but the most suitable research activities. The CeHRes Roadmap (Figure 1) is based on five principles: (1) eHealth development is a participatory development process; (2) eHealth development creates new infrastructures for improving health care, health, and well-being; (3) eHealth development is intertwined with implementation; (4) eHealth development is coupled with persuasive design; and (5) eHealth development requires continuous evaluation cycles [2,11]. The CeHRes Roadmap consists of 5 intertwined phases (the contextual inquiry, value specification, design, operationalization, and summative evaluation phase) that are connected by formative evaluation cycles [11]. The first 3 phases are focused on the development of eHealth. As the CeHRes Roadmap is very comprehensive, many of these principles are also important in other (aforementioned) eHealth development models. Consequently, although this review does not cover all eHealth development models, the identified research activities and lessons learned would also be suitable for application to other eHealth development models. To provide an overview of research activities used in studies guided by the CeHRes Roadmap, this narrative scoping review focuses on the following research questions:

1. Which research activities have been used in the development process of eHealth technologies that were based on the principles of the CeHRes Roadmap?
2. With which goal(s) and in which phase(s) have these research activities been used in the development process of eHealth technologies?
3. What are the experiences with, and lessons learned from, the use of the research activities?

Figure 1. The Center for eHealth Research Roadmap [2,11,18].



Methods

Inclusion and Exclusion Criteria

As the main goal of this study is to provide a focused overview of development activities used in the context of the CeHRes Roadmap and because no quality assessment of the to-be-included studies is required to reach this goal, a narrative scoping review was performed [19]. Records were included if they presented (part of) a development process of an eHealth technology that was based on the principles and/or phases of the CeHRes Roadmap. This refers to the use of ≥1 of the first 3 phases of the CeHRes Roadmap or explicit application of ≥1 of its principles. Consequently, studies were included if they referred to the original 2011 paper in which the CeHRes Roadmap was first introduced in either the introduction or methods section. This had to have been done in such a way that it became clear that ≥1 of the CeHRes Roadmap's phases or principles was used to inspire the design of the study. Studies that only referred to the CeHRes Roadmap in their discussion section and/or did not contain activities for eHealth development were thus excluded. Studies that only focused on implementation and summative evaluation were excluded. Records were also excluded if they did not present any data but merely discussed abstract guidelines or models for eHealth development. Furthermore, records not written in English, Dutch, German, or Portuguese were excluded. Finally, because of the broad scope and exploratory focus of this study, only study designs from peer-reviewed journals or books were included. Student reports, preprints, and poster abstracts were excluded because they were not peer reviewed.

Literature Search

To provide a complete overview of the studies that explicitly used or were inspired by principles of the CeHRes Roadmap, a straightforward search strategy was applied. Studies that referred to either the 2011 paper in which the CeHRes Roadmap and its principles were introduced or the new book chapter about the CeHRes Roadmap were identified in Scopus, Google Scholar, and Web of Science [2,11]. To include studies that were based on the principles of the CeHRes Roadmap but had been written before the 2011 paper, a snowball sampling strategy in which records that were coauthored by the founder of the CeHRes Roadmap (JEWC van Gemert-Pijnen) were searched in the same 3 databases. All searches were performed up until June 2021.

After removing duplicates in Covidence (Veritas Health Innovation Ltd), 3 researchers (HK, JK, and MCDS) screened the titles and abstracts using the aforementioned inclusion and exclusion criteria. As it might be possible that the development process was not fully explained in the title or abstract, the criteria were applied broadly to prevent the unjust exclusion of relevant articles. In case of doubt, a record was included to prevent overlooking relevant publications. Next, records were included for full-text screening if at least one of the authors decided to include an abstract. Full texts were assessed by 1 researcher (HK, MCDS, or JK) and, in case of doubt, discussed with one of the other researchers.

Data Extraction and Analysis

The data extraction process was performed by 3 researchers (HK, MCDS, and JK) and based on a table developed in an earlier study, which was used to present, and reflect on, eHealth development strategies [15]. All relevant information from the included records was copied into the data extraction table. The narrative data extraction form was divided into 3 main categories with accompanying subcategories and is presented in Multimedia Appendix 1. First, information on the overall goal and type of study design of the entire paper was included. Second, information was extracted for each development activity that was reported in the record. As in participatory or human-centered eHealth development processes, nonparticipatory activities such as literature reviews can also be valuable, no distinction was made between activities in which users were and were not actively involved [11]. In other words, *nonparticipatory* activities might also be valuable or even necessary for participatory development processes. For each activity, the following information was reported in the form: research goal, target group and participants, description of research activity used, rationale for research activity, main results that were obtained by means of the activity, and phase of the CeHRes Roadmap that the research activity was used in. If the phase of the CeHRes Roadmap was not explicitly mentioned in the record, this information was deduced by the authors using the goals and methods as reported in a recent publication on the CeHRes Roadmap [11]. Third, all lessons

learned about the application of the method that were reported in the records were copied into the data extraction form.

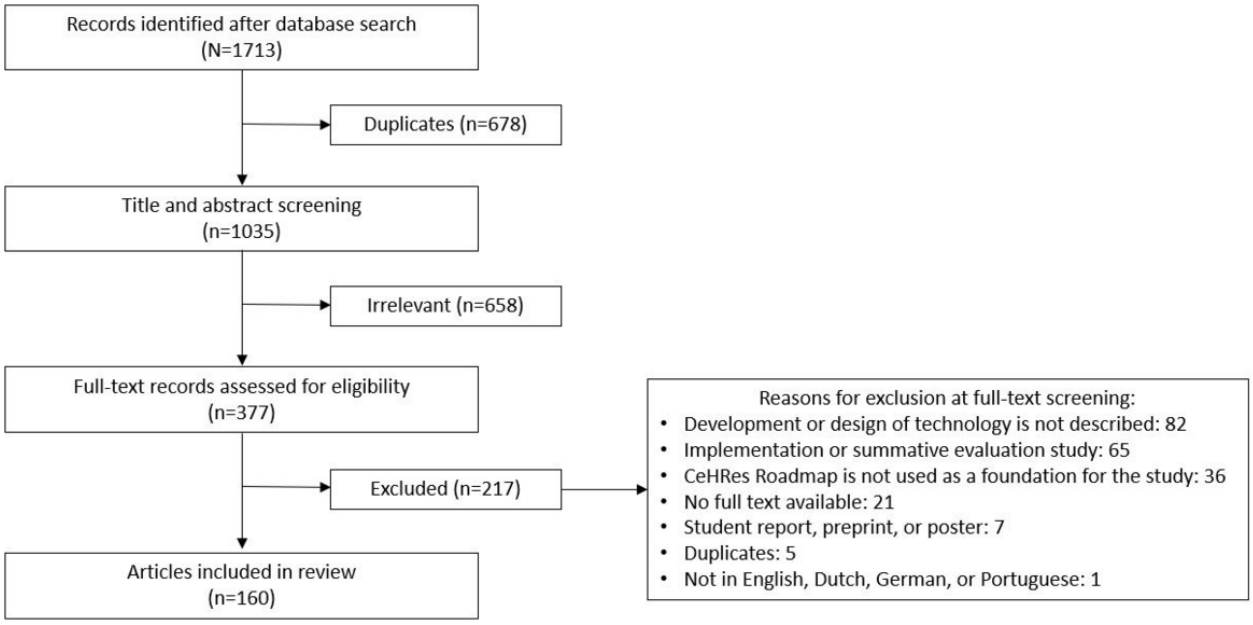
To analyze the data and answer the research questions, multiple steps were taken. To answer the first research question, an overview of research activities used in all studies was created. As activities were often named in slightly different ways, researchers formulated overarching categories for development activities by means of discussions until consensus was reached. In addition, a definition for each research activity was formulated. This definition was created by means of the information provided by the authors of the included records. If necessary, the definition was subsequently fine-tuned. This was done using other relevant literature—mainly a book that was edited by the research group of this paper's authors [5]—and discussion among the authors of this paper. To answer the second research question, all information on the goal of a research activity, its main results, and the phase of the CeHRes Roadmap in which it was used was combined into 1 document. Researchers used this information to summarize the ways in which an activity was used and identify examples to illustrate the goals that can be achieved with the research activity. Again, if necessary, discussion among the authors took place until consensus was reached. Third, to answer the final research question on lessons learned, all extracted fragments with information about lessons learned were provided per activity; 2 researchers (HK and JK) went through these fragments separately and individually summarized the most important lessons learned. After discussion, an overview of all lessons learned was created. On the basis of this overview, the 3 most prevalent and applicable lessons learned were selected by the researchers and presented in a narrative way.

Results

Search Results

As can be seen in Figure 2, the initial literature search yielded 1713 unique records. After title and abstract screening by 3 researchers (HK, JK, and MCDS) using the aforementioned inclusion and exclusion criteria, of the 1713 records, 377 (22.01%) remained. After full-text screening of these 377 records, 160 (42.4%) were included. The main reason for excluding full texts was a lack of specific focus on the development of an eHealth technology.

Figure 2. Search strategy and results. CeHRes: Center for eHealth Research.



Study Designs and Technologies

An overview was generated of the designs of all included studies (n=160). Most development studies used a multi-method approach, in which various qualitative and quantitative activities were combined (90/160, 56.3%). Other designs that were identified were qualitative cross-sectional (21/160, 13.1%), quantitative cross-sectional (18/160, 11.3%), literature studies (17/160, 10.6%), qualitative longitudinal (12/160, 7.5%), and quantitative longitudinal (2/160, 1.3%).

The included studies focused on a broad range of eHealth technologies. The main goal and described methods per study are provided in Multimedia Appendix 2 [8,15,20-176]. Most of the studies focused on web-based interventions (74/160, 46.3%) such as a web-based module for treatment of depression or a decision support tool for nurses. Mobile apps were the focus of development in 21.3% (34/160) of the studies. Apps were used, among other things, to support patients with breast cancer in doing arm and shoulder exercises and to support citizens in dealing with tick bites. In addition, virtual reality (VR) was studied in 1.9% (3/160) of the papers, which mostly focused on role-playing in treatment of patients with psychiatric disorders. Furthermore, in 16.9% (27/160) of the studies, there was no clear description of the technology, either because it was too early in the development process or because the authors did not include a description.

Overview of Methods and Products Per Phase of the CeHRes Roadmap

During the data extraction process, a distinction had to be made between research methods and products. Methods such as interviews were used to collect new data, and products such as prototypes were based on, or summarize, previously collected data and can be used as tools to collect new data. In Table 1, an overview of all methods and products that were identified is provided, including references to the accompanying studies (Multimedia Appendix 2). In addition, for each method or product, the number of studies that used it in the contextual inquiry, value specification, or design phase is provided. As in some studies multiple methods or products were used in different phases or because methods that were relevant for >1 phase were used, the sum of the columns is not the same as the number of studies. In 30% (48/160) of the studies, the phases of the CeHRes Roadmap were mentioned explicitly; in the other studies, the CeHRes Roadmap was mostly used to inspire the design. Two authors (HK and JK) categorized the methods and products in these papers based on the definitions of the phases of the CeHRes Roadmap [11].

Table 1. An overview of the identified methods and products, the accompanying references, and the phases of the CeHRes (Center for eHealth Research) Roadmap in which they were categorized in the included studies (N=160).

Method or product[a]	CeHRes Roadmap phases			References
	Contextual inquiry, n (%)	Value specification, n (%)	Design, n (%)	
Methods				
Focus group (n=71)	45 (63)	62 (87)	29 (41)	[8,15,20-95,177]
Interview (n=70)	41 (59)	32 (46)	21 (30)	[15,21,24,26,28-31,33,34,40,44,45,47,49,50,52,54-56,58,61,67,69,70,74,76,78,79,81,83,84,88,89,91,96-129,178]
Questionnaire (n=51)	28 (55)	16 (31)	19 (37)	[15,24,26,29,32-34,46,49,56,60,67,69,71,76,80,84,89,96-98,101,104,107,109,115,125-127,129-151]
Usability test (n=51)	0 (0)	1 (2)	67 (131)	[8,20,21,24,25,37-39,41-43,45,48-50,54,55,63,64,66,67,71,72,75,77,80,81,84,92,98,102,104,106,107,109,110,112,113,117,118,127,131,152-156]
Literature study (n=43)	36 (84)	2 (5)	4 (9)	[8,20,21,27,28,38,43,46,53,55,58,69,72,75,80,81,84,92,93,99,104,112-114,129,153,157-173]
Desk research (n=15)	14 (93)	3 (20)	0 (0)	[15,31,33,34,43,51,52,56,59,61,88,91,98,112,135]
Log data analysis (n=10)	1 (10)	0 (0)	8 (80)	[68,86,104,108,111,133,140,174-176]
Card sorting (n=9)	1 (11)	1 (11)	8 (89)	[8,25,34,36,70,77,98,102,125]
Delphi study (n=3)	1 (33)	2 (67)	0 (0)	[72,84,101]
Experience sampling (n=1)	1 (100)	0 (0)	0 (0)	[130]
Products				
Prototype (n=32)	0 (0)	8 (25)	29 (91)	[15,25,31,36,40,46,48,51,52,55,63,66,69,72,74,75,77,78,81,83-85,88,92,94,96,102,107,110,113,115,121,124,127,129]
Requirements (n=11)	0 (0)	10 (91)	0 (0)	[20,30,31,38,39,70,91,92,114,115,129]
Stakeholder map (n=10)	4 (40)	4 (40)	0 (0)	[15,27,30,35,46,56,59,129,135,153]
Values (n=7)	0 (0)	7 (100)	0 (0)	[15,79,81,97,109,129,152]
Behavior change strategies (n=5)	0 (0)	1 (20)	4 (80)	[69,79,84,91,109]
Personas (n=5)	—[b]	1 (20)	—	[67,70,73,74,81]
Business model (n=4)	0 (0)	4 (100)	0 (0)	[27,62,84,99]

[a]The sum of the times a method was used in the contextual inquiry, value specification, or design phase is higher than the number of included studies per method because in multiple studies, one method was used more than once in the development process.

[b]No relevant records were identified for the category.

Definition, Applicability, and Lessons Learned Per Method

Overview

In the following sections, the definition that was generated by means of the included studies is provided for each method. In addition, different examples of how the method was used are given. Finally, the 3 most relevant lessons learned that were mentioned in the included papers are summarized.

Focus Group

Definition

Focus groups refer to meetings where qualitative data are collected by involving a relatively small number of stakeholders in a group discussion. This discussion is focused on a particular topic or set of issues, ranging from relatively unstructured *workshops* and generative design sessions to highly structured meetings.

Examples of Applications

Of the 160 included studies, 71 (44.4%) were focus group studies, most of which had some sort of predetermined structure. The extent to which data were systematically analyzed differed among the studies. In some, extensive coding schemes were created, whereas in others, the most important findings were summarized. Furthermore, some focus groups included a single type of stakeholder, for example, only patients, whereas others included a combination of multiple stakeholders such as therapists, patients, and technology developers. In addition, some studies included novel participants in each focus group, whereas others used recurring *coresearchers* [90]. As can be seen in Table 1, focus groups were used throughout all phases of the development process. This shows that focus groups can be used to reach a broad range of goals. Examples of these goals include the following: to identify points of improvements of the current situation, such as self-management of patients with chronic obstructive pulmonary disease (COPD) [32] or care for cerebral palsy [33]; to discuss the possibilities of a specific technology, such as the values of people with obesity regarding a to-be-developed behavior change intervention [79]; to gain insight into cognitions such as attitudes toward measures to reduce antimicrobial resistance [8]; to identify or validate values or requirements with potential end users such as health care providers [23,93]; or to collect input for the improvement of a prototype, for example, for a portal for infection control [27].

Lessons Learned

First, authors of multiple studies indicated that focus groups are a good way to gain more insight into the specific needs, wishes, and opinions of individuals regarding eHealth. To achieve this, focus groups can be conducted with a group of similar or very different stakeholders. Participants can bounce ideas off of each other and can directly respond to each other and can provide insight into the prevailing consensus or the range of different opinions or perspectives regarding eHealth [35-37,43,54,57,58]. However, researchers should take potential power imbalances or potentially sensitive conflicting values into account when inviting participants of a focus group. A second lesson refers to the iterative nature of eHealth development. Focus groups can be used in a *sequential* way: multiple focus groups can be conducted in a row and the goal and content of each focus group can be based on the outcomes of the previous focus groups. However, this iterative approach was said to be quite time consuming for researchers and participants [33,54,64,82,84-86,90,92]. Third, to ensure that valuable information for eHealth development is gathered, the content and form of focus groups need to be adapted, based on the topic and target group [39-41,51,55,61,87]. To illustrate, in-person focus groups are not suitable for every topic. Web-based alternatives might be considered when, for example, sensitive topics such as sexual health are discussed. Furthermore, different types of participants might require different types of focus groups. For example, focus groups with people with an intellectual disability or with older adults require a setup with more concrete examples of eHealth and might benefit from *icebreakers* and room for informal conversations [87]. In contrast, focus groups with therapists or researchers can cover more abstract topics [15].

Interview

Definition

In interviews, individuals are asked questions in a structured, semistructured, or unstructured way to obtain answers from a broad range of possible stakeholders, guided by an interview scheme.

Examples of Applications

Of the 160 included studies, 70 (43.8%) featured interviews that took place at multiple points in the development process. Interviews can be conducted from the start of a development process to not only analyze a problem, but also evaluate prototypes. Consequently, interviews can have a broad range of goals. Examples include identifying points of improvement for a current situation such as treatment of forensic psychiatric patients [96]; analyzing target or risk groups in, for example, tick bites [100]; identifying points of improvement for prototypes or existing (eHealth) interventions according to end users or design experts [83,84]; identifying potential barriers and facilitators for implementation later in the process, such as high costs and required skills training [78,96]; describing a current behavior and its determinants [91]; collecting experiences of participants after letting them try out an app in real life [83]; and generating or validating values and requirements [15,28,98,103].

Lessons Learned

First of all, in multiple papers, authors mentioned the importance of individual, in-depth interviews to incorporate the perspective of vulnerable, complex target groups such as people with dementia or severe mental illness in eHealth development [15,21,61,70,96,97,109,111,121,123,127]. This is especially important because perspectives of these underserved target groups are often overlooked in eHealth development. However, despite the benefits, including these types of target groups was found to be challenging, mostly because participating in relatively long, in-depth interviews requires a fairly high level of cognitive abilities such as attention and memory. Consequently, researchers should account for the characteristics

of their target groups by, for instance, keeping the interviews as short as possible [78,128]. Another option is the use of concrete examples of eHealth technologies to account for response or recollection biases [24,52,97,121,123]. Second, although interviews can yield valuable results, a limitation is that they can offer a 1-sided picture of stakeholders' needs and wishes regarding eHealth. Selection bias can result in a sample that is overly positive or negative [78,128]. To overcome these issues, which are related to generalizability, multiple authors recommended that interviews be combined with other methods in a multi-method or mixed methods approach [24,40,52,54,69,78,96,97,100,102,108,109,116,118,120,124,126]. This could be done by combining interviews with a small sample size with a questionnaire with a larger sample size. Although small sample sizes were not necessarily considered problematic in eHealth development, combining methods in an iterative way was suggested as a way to overcome issues with generalizability [81,127]. Third, interviews were used quite often and were generally viewed as a useful method that can be used at any point in the eHealth development process [15,21,45,54,96-98,100,105,106,109].

Questionnaire

Definition

A questionnaire can be either quantitative or qualitative; it consists of a series of open- or closed-ended questions for the purpose of gathering information from—often—a relatively large sample of respondents and can be distributed on the web or on paper.

Examples of Applications

Of the 160 included studies, 51 (31.9%) featured questionnaires that were applied in all phases of the development process, which means that they can be used for a broad range of goals. Examples include gathering information for stakeholder identification and analyses [135]; identifying points of improvements in conceptions and knowledge of stakeholders on infection outbreaks and antimicrobial resistance [134,178]; mapping attitudes toward technologies such as embodied conversational agents [132]; identifying needs and wishes (values) regarding a to-be-developed technology [97]; or evaluating low-fidelity prototypes, for example, scenarios on multiple possible VR interventions [15]. In questionnaires, either new questions can be generated by researchers, based on previous research, or existing questionnaires can be used, for example, the Personal Involvement Inventory, the eHealth Literacy Scale, or System Usability Scale [15,24,84,127].

Lessons Learned

First, in many studies, authors reflected on possible biases that might arise when using questionnaires to develop eHealth [24,69,80,97,109,115,126,134,138,139,142,146,149]. Among other things, vulnerable target groups with low literacy skills and no internet access were often hard to include in questionnaires [96]. Furthermore, some researchers used students or services such as Amazon Mechanical Turk to generate large samples; however, this raises questions about the generalizability of the results [136]. This means that results of questionnaires have to be interpreted with care and should not serve as the sole input for an eHealth technology. Second, multiple authors indicated that a questionnaire, especially one on the web, is a suitable method to quickly and efficiently collect data from different types of stakeholders and to check for differences among groups in, for example, opinions or knowledge [60,67,69,98,126,131,147,150]. A pitfall of this approach is that collecting rich in-depth information about, for example, an existing problem or a prototype is challenging. The main reason is that participants often do not provide elaborate answers to open-ended questions; in addition, it is not possible to ask probing questions [15,139,141,146]. The third lesson learned is that questionnaires need to be combined with other types of data such as interviews or focus groups to meaningfully contribute to the development process [24,26,33,60,71,96,97,126,150]. As eHealth development requires a complete picture of the current situation and needs and wishes of the stakeholders, triangulation of methods should be used. For example, products generated earlier, such as values, can be cross-referenced or interviews can be used to provide more context to the outcomes of a questionnaire.

Usability Testing Methods

Definition

Usability testing is an umbrella term that can refer to a broad range of methods such as a think-aloud method with scenarios, cognitive walkthrough, heuristic evaluation, or eye tracking. These methods are used to conduct formative evaluations of prototypes by testing them with participants such as potential users or experts. Usability refers to the extent to which a user can use a product effectively and without effort, immediately learning its use. Usability tests can be used to identify usability problems, flaws, and points of improvement or gather overall opinions.

Examples of Applications

Of the 160 included studies, 51 (31.9%) reported on usability tests. As these tests require a prototype that should be based on earlier research, they are often not conducted at the beginning of the development process when the scope and content of the to-be-developed eHealth technology still have to be determined. An exception is when an existing technology is evaluated in the contextual inquiry to collect input for redesign. Possible goals of usability testing are to identify points of improvements of a low- or high-fidelity prototype according to experts and/or end users (see the *Prototype* section under *Definition, Applicability, and Lessons Learned per Product*) [77,81,127,152]; to evaluate a technology's potential to improve problems in a specific organization [77]; to analyze the way a high-fidelity prototype is used by prospective end users [24,84]; to assess whether the prototype fits the current work practice of end users such as nurses [25]; or to generate new, or further specify, values or requirements [131].

Lessons Learned

First, multiple authors indicated that usability tests should be conducted with a broad range of stakeholders: not only end users such as patients, but also caregivers, managers, technology developers, and experts on content and design [21,43,54,64,77,112,127,131,152,155]. Different types of

participants can provide different kinds of feedback on an eHealth technology. It was suggested that experts on, for example, usability or persuasive design can be included by means of cognitive walkthroughs or heuristic evaluations, whereas users can be involved through think-aloud procedures, guided by scenarios [127]. Second, authors stated that ideally, multiple methods should be combined to paint a full picture of a prototype's usability [24,25,64,77,80,118,152]; for example, qualitative methods such as think-aloud procedures and interviews can be combined. Qualitative approaches can also be combined with quantitative data collected by means of, for example, log data analyses, questionnaire data, or eye tracking [127]. A way to do this is by using an iterative approach based on the user-centered design framework [81,127]. Third, in multiple studies, authors indicated that values or requirements of the to-be-developed eHealth technology can be used to guide usability tests. These can be used to structure data collection by, for example, verifying whether the requirements are present in the technology, or to analyze the data by means of deductive coding using the values to ensure that everything is well aligned [8,21,42,54,113,131].

Literature Study

Definition

Although there are many ways to conduct a literature study, for example, rapid, systematic, or scoping reviews, they all aim to create an overview of a certain topic using scientific literature, often in a systematic manner.

Examples of Applications

Of the 160 included articles, 43 (26.9%) were literature studies, most of which were conducted at the start of a development process to create an overview or get acquainted with a specific topic. The included studies in this review ranged from relatively unstructured, quick literature scans to elaborate systematic reviews in line with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. An example of a literature study is a systematic review, which is often used to provide a complete, exhaustive summary of the current literature on a specific topic in a highly structured way. Another example that is often used in eHealth research is a scoping review, which offers insight into the status quo of scientific literature in a certain broad field of study by means of a systematic search of literature without paying too much attention to the quality of the studies. Possible goals of literature studies include the following: to gain insight into specific problems such as antimicrobial resistance or into broad domains such as technology in forensic psychiatry [158], to gain insight into development-related questions such as suitable persuasive features or behavior change techniques (BCTs) [159], or to develop recommendations for the design of specific interventions such as embodied conversational agents [172].

Lessons Learned

First, in several eHealth development studies, literature reviews were conducted quite rapidly and a systematic approach was not described. However, multiple authors indicated that it is always important to use a systematic approach when planning and executing any sort of literature review to ensure completeness of the results [158,159,166,167,173]. Second, multiple authors pointed out that often a lot of time has passed between initial data collection and publication of studies—which is especially relevant for the rapidly changing domain of eHealth. Consequently, not all state-of-the-art knowledge on technologies is published at the point of conducting the review [99,158,163,167,172]. Therefore, researchers can combine scientific literature with gray literature such as policy documents, company reports, or communication about work in progress to provide a more comprehensive overview of the current state of affairs. Third, because many literature studies in the contextual inquiry have a broad setup, it is often not possible or desirable to include a quality appraisal or only include experimental studies. If this is done, much valuable information on, for example, potential applications of a technology might be missed. In any case, it is important to reflect on the choices that were made in the reviewing process when interpreting the results [158-160].

Desk Research

Definition

Desk research refers to the nonsystematic collection of often nonscientific material such as presentations of an intervention, management reports, project documents, or activities or tasks of stakeholders. This material supports the development team in learning as much as possible about a specific topic.

Examples of Applications

Of the 160 included studies, 15 (9.4%) were desk research studies, which ranged from very unstructured studies, for example, talking to people or checking correspondence, to relatively structured studies, for example, a systematic analysis of the content of nonscientific documents or protocols. Desk research can be used throughout all phases of the development process when questions arise that do not require thorough, replicable research. In the included studies, desk research was used most often at the beginning of a development process to gain insight into specific fields or problems. Examples of applications include the following: to identify stakeholders, to gather information on protocols or guidelines [34,98], to create an overview of nonpublished projects on a technology in a specific sector [15], to observe existing face-to-face workshops to gain insight into their content [66], to gather correspondence for the analysis of existing communication processes [33], or to search for similar (eHealth) interventions [43].

Lessons Learned

First of all, authors stated that desk research should not be viewed as a synonym for randomly collecting information. Desk research should be used as a method that requires a clear research question and, if possible, a systematic search strategy that is clearly connected to the goal of the eHealth development process [15]. To illustrate, in the study by Breeman et al [81], desk research that was used to evaluate existing eHealth apps was structured by means of a newly developed evaluation tool for apps. Second, there is not necessarily a limit to the kinds of materials that can be included in desk research, depending on the research question. Some examples are apps that are available in an app store, scientific literature, protocols on cognitive

behavioral therapy, written communications, policy documents, and presentations at meetings or conferences [33,98]. Third, it was suggested that desk research is a suitable method to look outside of a project's scope to learn from the use of an eHealth technology in other domains. For example, researchers in mental health care investigated how VR is used in other settings such as hospitals or advertising [15].

Log Data Analysis

Definition

Log data are objective registrations of events that can be recorded on an individual basis, such as logging in to a website, entering a room (measured by sensors), sending a message, or performing a certain action in an app. These often large data sets can be analyzed to gain insight into behavior patterns.

Examples of Applications

Although log data are often used for the evaluation of the use of an already implemented eHealth technology, of the 160 included studies, 10 (6.3%) were log data studies that showed that log data can also provide valuable input for the development process. Log data can be analyzed in different ways, ranging from descriptive statistics to more complex analyses such as Markov modeling. In development, log data can, for example, be used to gain insight into use and points of improvement of (high-fidelity) prototypes [86,175] or to provide insight into dos and don'ts for the design of similar to-be-developed systems by analyzing existing similar websites [176].

Lessons Learned

First, authors stated that log data can be used in many ways to gain insight into online or offline behavior of prospective end users, not just eHealth use. Analyses can be performed with *real-life* log data on, for example, the use of an existing eHealth technology or with log data on the use of interaction-enabled prototypes in laboratory settings. Log data can also be collected on behavior in real-life settings that is not related to the use of an intervention but is useful for the contextual inquiry. An example is logging the number of times that doors within a nursing ward were opened [111,175]. Second, when log data on a prototype or newly developed eHealth technology are collected, researchers have to discuss with designers what data they want and how they should be collected beforehand. A well–thought out activity log protocol that describes which features should be logged should guide this debate [175]. Third, although very complicated analyses can be conducted with log data, straightforward descriptive statistics can often also be very valuable for researchers during the eHealth development process. This makes log data analysis a less complicated method than is often assumed [175].

Card Sorting

Definition

Card sorting is a method that can help design or evaluate the information architecture and structure of a technology. It can be used to structure units of information provided on cards. A distinction can be made between closed- and open-ended card sorts. In closed-ended card sorts, the main categories are provided by the researchers, whereas in open-ended card sorts, participants create their own main categories.

Examples of Applications

Of the 160 included studies, 9 (5.6%) involved card sorting. Card sorting is often used a bit further along the development process because it is often focused on the content and structure of an eHealth technology. Card sorting can be used to create new information structures or validate existing ones, for example, the structure of menus on websites with infection control guidelines [77,98], or to create a logical content structure that is in line with the users' mental model, for example, for apps or web-based tools for health care workers [34].

Lessons Learned

Multiple authors indicated that card sorting is a fast, efficient, and cheap method to gain insight into the structure of a to-be-developed eHealth technology, especially when conducted on the web [25,70,77,98]. Web-based card sorting using programs such as Optimal Workshop software was said to be less prone to error, less labor intensive, and more efficient because larger numbers of participants can be included with less effort. Second, card sorting offers a concrete way to support people in expressing their thoughts and needs regarding an eHealth technology. This helps researchers gain more insight into how information should be structured. Third, to paint a complete picture of prospective end users' requirements for the eHealth technology, card sorting should be combined with other methods such as interviews or focus groups [25,77,98,102]. In addition, the categories or structures that result from card sorting should be validated by using methods such as usability testing or interviews [25].

Delphi Study

Definition

A Delphi study offers a systematic way to determine consensus among various stakeholders, mostly experts. It solicits opinions from groups in an iterative process of answering questions, usually in multiple rounds.

Examples of Applications

Delphi studies can be used at multiple points in the eHealth development process when consensus among stakeholders—often experts—is needed. They can, for example, be used to gain insight into preferences for the eHealth technology or to reach consensus about values or requirements. Of the 160 included studies, 3 (1.9%) were Delphi studies, which used this method to identify expert recommendations for parents to reduce the risk of depression or anxiety in their children to include in the eHealth technology [72], to reach consensus on the most optimal way to integrate an eHealth technology in standard treatment [101], and to gain insight into expert opinions on relevant self-management behaviors for reducing the impact of COPD [84].

Lessons Learned

First of all, authors explained that Delphi studies have to consist of multiple rounds of data collection to reach consensus. This can be done using web-based questionnaires or focus groups in which participants have to participate at least two times [72].

Other types of data can be used as a starting point, such as scientific literature or outcomes from interview studies. Second, it can be challenging to recruit participants for a Delphi study because, for example, there are not many experts in a specific or new field, which can be especially relevant when studying novel applications of an eHealth technology [101]. This implies that researchers have to carefully think about whom to involve, for example, through snowball sampling, before starting with the study. Third, reaching consensus can take time and can be complex. This means that Delphi studies can require a lot of time from researchers and participants, which is not always available in eHealth development processes [101].

Experience Sampling

Definition

Experience sampling—sometimes also referred to as ecological momentary assessment (EMA)—is a structured *diary method*. It can be used to gather relevant, subjective experiences such as physical symptoms, mood, and behavior in daily life using multiple measures throughout the day using pen and paper or technology such as apps or wearables.

Examples of Applications

Of the 160 included studies, only 1 (0.6%) involved experience sampling, which was used to gain insight into information-seeking behavior during an EHEC (enterohemorrhagic *Escherichia coli*) outbreak [130].

Lessons Learned

Although experience sampling was used in only a single study, multiple lessons learned were provided. These lessons are not directly related to, but seem to be very valuable for, eHealth development. First, experience sampling was found to be a suitable method to explore an existing situation and thus to provide valuable input for further eHealth development [130]. Second, including participants who are willing and able to constantly provide input on their experience for a longer period of time can be challenging and time consuming, which also applies when it comes to ensuring that participants do not drop out during the study [130]. Third, when researchers want to study experiences during a specific event such as a virus outbreak, they ideally want to start collecting data from the beginning of the outbreak. However, in practice, this is often impossible. This can be compensated for by using a survey with retrospective questions; however, participants might be unable to correctly and completely recall their experiences [130].

Definition, Applicability, and Lessons Learned Per Product

Overview

Whereas methods are used to collect new data, products are based on, or summarize, these previously collected data. They can also be used as tools to collect new data. For each identified product, the definition, ways of applying it, and a maximum of 3 lessons learned are described in the following sections.

Prototype

Definition

Prototypes are visual representations of a to-be-developed technology, ranging from low- to high-fidelity representations. Low-fidelity prototypes often do not contain much detail, allow no automatic interaction between user and prototype, and can be relatively easy to create. Examples are paper-based sketches or wireframes, possibly combined with scenarios. High-fidelity prototypes are mostly digital; often cost more time, money, and technical skills to develop; and allow for user interactions, such as programmed apps or digital, interaction-enabled prototypes.

Examples of Applications

Prototypes were used in 20% (32/160) of the included studies, and in these studies, most prototypes were created during the design phase. However, it is also possible to create low-fidelity prototypes early in the development process, for example, when presenting initial ideas about a technology to participants [15]. Furthermore, end users or other stakeholders can create prototypes themselves in cocreation sessions to visualize their ideas and preferences [113]. Prototypes are often based on values and requirements [15,55,113]. In addition, multiple prototypes are often created and improved based on outcomes of usability testing [52,88].

Lessons Learned

First, the included studies showed that there is not 1 single way to create a prototype. Methods of creating a prototype can range from very *quick and dirty* without any content to very complex, creating a highly interactive product [40,55,121]. This depends on the goal, for example, to check the overall structure of a prototype or to identify usability problems. Second, it is important to have a rough idea of the costs of developing a technology as soon as possible, mostly to prevent the final version of the prototype from seeming to be too expensive. Inclusion of technology developers from the start of the development process was recommended to prevent these problems [46,113]. Third, because major changes might be made to prototypes in an iterative development process, changes to the requirements might also be necessary to ensure that they remain in line with the prototype of the eHealth technology. This highlights the importance of an iterative approach [55,113,115]. In line with this, making major changes to prototypes could require time and resources that are not available, which might result in a suboptimal prototype [88].

Requirements

Definition

Requirements are short statements that prescribe what is *required* of a technology: “They describe what a technology should do, what data it should store or retrieve, what content it should display, and what kind of user experience it should provide” [179].

Examples of Applications

Requirements were formulated in 6.9% (11/160) of the included studies. They were never formulated at the beginning of the development process because they should be based on outcomes of earlier activities. The included studies showed that

requirements can be based on data generated by 1 method, such as interviews, or on a combination of data from different methods and scientific literature [91]. Requirements can be based on values, where values serve as a *bridge* between the previously conducted research and the specific requirements [129]. Requirements can be used to specify previously formulated values, to serve as foundations for prototypes, or to communicate needs and wishes to developers and discuss these needs and wishes with them [15,30].

Lessons Learned

First of all, authors stated that requirements should be elicited in a systematic way. Ideally, multiple sources of data should be combined, such as scientific literature and qualitative data collected from multiple types of stakeholders [91,114,115]. When combining sources, development teams need to be aware of conflicting requirements, for example, a discrepancy between user needs and scientific literature [91,92]. Second, it might not be possible to include all requirements in an eHealth technology because of practical, technical, or financial limitations. However, that does not mean that it is not worth the effort to further specify them: these requirements might be incorporated in the technology at a later point in time [38,39]. Third, it was mentioned that eliciting requirements from target groups comprising patients who are vulnerable and clinically complex, such as people with psychosis or dementia, might be challenging. Consequently, researchers should carefully select methods that fit the characteristics and skills of these populations [38,39,70].

Stakeholder Map

Definition

A stakeholder map is a visualized overview of stakeholders—people or organizations who affect or are affected by an eHealth technology—and their interrelationships.

Examples of Applications

Stakeholder maps were reported in 6.3% (10/160) of the included studies. Stakeholders should be identified from the start of the development process, and this overview should be updated throughout the entire process. Examples of stakeholders are patients, experts on a specific topic such as depression, a commercial company that can develop apps or VR, health care providers, researchers on eHealth or the health problem at hand, and employees of other organizations that could use the intervention later [15,129,153]. The roles and tasks of these stakeholders regarding the to-be-developed eHealth technology should be identified by means of methods such as interviews or desk research.

Lessons Learned

First, in the included papers it was shown that stakeholder maps are not created from scratch but are based on data. Often, a stakeholder map is created by means of stakeholder identification: the systematic process of finding out who the stakeholders of an eHealth development process are. Stakeholder identification can be supplemented by a stakeholder analysis, which refers to the analysis of interdependencies, responsibilities, and stakes of the identified stakeholders. To achieve this, ideally, a combination of research methods such as questionnaires, interviews, or literature reviews is used [27,129]. Second, the studies stated that the importance of a stakeholder map should not be underestimated. Thorough investigation of stakeholders and their context is very important not only for the entire development process, but also for implementation and evaluation of the to-be-developed eHealth technology [30,35]. Consequently, the stakeholder map should be constantly updated throughout the process because new insights might arise or new stakeholders might emerge [35,59]. Third, it was considered to be important to include a wide variety of stakeholders, where researchers should look beyond their own setting. For example, when developing a VR intervention for forensic mental health care, researchers should also include stakeholders in other organizations where VR is used, such as hospitals [15,35,153].

Values

Definition

Values refer to ideals or interests of stakeholders: they specify what stakeholders want to achieve or improve by means of an eHealth technology and capture what the added value of a technology should be for the people and organization involved.

Examples of Applications

Of the 160 included studies, 7 (4.4%) involved values, which were created based on outcomes of previously conducted research. This means that values are often formulated later in the development process, mostly during the CeHRes Roadmap's value specification phase, hence the name. Values remain relevant throughout the remainder of the process because they can also serve as foundations for requirements and prototypes. Values can be used to summarize or synthesize outcomes of previously conducted studies such as interviews or questionnaires. They can also serve as foundations for prototypes, implementation plans, or evaluation goals [97]. Examples of values are *improvement of skills*, *easy to use in current treatment*, *affordability*, *self-management*, and *positive self-image* [79,97].

Lessons Learned

First, authors stated that values should be formulated in such a way that they are neither too specific to prevent overlap with requirements nor too broad and vague. A shared understanding among the eHealth development team members about what values are is essential to achieve this [15,64,97,152]. Second, values should capture the whole range of stakeholder needs and wishes, not merely those of end users of the eHealth technology [81,129]. To paint a complete picture, multiple methods should be used. In addition, when new insights arise, values should be updated by a multidisciplinary research team to ensure that they continue to align with the perspectives of the key stakeholders [15,62,79,97,129]. Third, conflicting values might arise related to, for example, costs or the focus of an eHealth technology. A good way to resolve these conflicts is by discussing them with a group consisting of multiple types of stakeholders [79,97]. In addition, researchers should not view a value map as static: it might have to be adjusted based on changes in the context and users [129].

Behavior Change Strategies

Definition

Behavior change strategies such as evidence-based BCTs or persuasive elements can be integrated in the design of eHealth technologies to increase their effectiveness.

Examples of Applications

Of the 160 included studies, 5 (3.1%) featured behavior change strategies, which were mostly used in the design phase. The product was used to determine which theory-based methods should be included in an intervention to increase the chances of achieving behavior change. Examples are the inclusion of BCTs such as goal setting and self-monitoring in a mobile app to increase vegetable consumption [69] or the application of 3 theory-based methods and 4 accompanying strategies to influence the attitudes and skills of patients to support them in their communication with health care professionals [109]. In the study by Asbjørnsen et al [79], researchers used methods from design thinking to translate values and needs of people with obesity into persuasive features and behavior change theories such as goals and planning, personalization, and self-monitoring.

Lessons Learned

First, behavior change theories should ideally be combined with outcomes of human-centered design methods, instead of being mostly researcher based. However, more insight is required into how this should be done in eHealth development [79,109]. Second, operationalizing theoretical strategies into a user-friendly eHealth technology might be challenging [79,91]. To overcome this, participatory approaches such as cocreation sessions or prototyping workshops can be used to determine how theoretical working mechanisms can be translated into features of an eHealth technology [69,79]. Third, traditional behavior change theories might be too static to integrate into adaptive eHealth technologies, especially in the case of just-in-time personalized interventions. This implies that new types of theories might be required [69].

Persona

Definition

Personas are *user archetypes* that summarize a representative person from the target group. They consist of a description of different types of characteristics of a future or actual user, often in the form of a story.

Examples of Applications

Of the 160 included studies, 5 (3.1%) described personas. In the study by Dick et al [73], 3 personas of users of an eHealth intervention for illicit substance use were created: the heavy user, abstainer, and occasional user, whereas the study by Derks et al [70] developed personas for an intervention for people with cardiovascular disease and the study by Breeman et al [81] developed personas for an intervention for people with borderline personality disorder.

Lessons Learned

First of all, it is recommended to use existing guidelines and frameworks such as that of LeRouge et al [180] when developing personas for eHealth technologies. These frameworks should be combined with human-centered research methods such as focus groups [67,70,81]. Second, to structure the persona-building process, researchers need to identify characteristic categories that need to be included in the persona, in which attention should also be paid to skills and attitudes related to the to-be-developed eHealth technology. Examples are demographics and personality of the service user; their medical and psychological profile, including fears and motivations for behavior; their abilities, (technological) skills, and coping strategies; and their needs and goals [67,70,73,81]. Finally, personas were seen as a useful tool to tailor the content and design of eHealth technologies and can be connected to the requirements or to BCTs [73].

Business Model

Definition

A business model captures how an organization creates, delivers, and captures values; it describes how an organization conducts its business. It is a conceptual and analytical framework to map, discuss, and help realize the added value of an eHealth technology, as well as to determine the key factors that are associated with a sound and sustainable implementation.

Examples of Applications

Of the 160 included studies, 4 (3%) described business models. Often, the development of business models is initiated during the first stages of a development process because development and implementation should be intertwined [84]. However, the studies showed that a business model is not finished during development: it should be updated throughout the entire development and implementation process. An example of its application is the use of the business model canvas for an eHealth portal for infection control. This model includes the technology's key partners, key activities and key resources, cost structure, revenue streams, value proposition, customer relationships, customer segments, and channels [27].

Lessons Learned

First, researchers concluded that perspectives from all important stakeholders should be accounted for in a business model. This can be done by means of focus groups with multiple stakeholders, integration of earlier collected data, or in-depth interviews [84]. When collecting input for the business model, it was indicated that questions to participants should be very concrete because abstract questions will yield equally abstract and thus less useful answers [99]. Second, experience has shown that creating a business model is very time consuming, partly because there are no business models specifically for eHealth yet [27,99]. Third, a business model often does not have a fixed end and needs to be adapted continuously throughout the implementation and evaluation phases of an eHealth technology [27].

Discussion

Principal Findings

The main goal of this narrative scoping review is to create an initial overview of the methods used in eHealth development

processes guided by the CeHRes Roadmap. Furthermore, we aim to identify for what purposes and in which phases these methods are used and provide an overview of the most relevant lessons learned. During the analysis, it became clear that a distinction between development methods and products can be made. In the 160 included studies, 10 overarching methods and 7 products were identified. Most of the identified methods were used in all 3 development phases of the CeHRes Roadmap. They were used for a broad range of goals, underlining the many different possibilities that exist for eHealth development. The lessons learned showed that most authors agreed that the methods and products contributed to eHealth development in a positive way by providing more insight into the users and context. However, a critical reflection on the methods or products and accompanying conclusions related to a method or product not being suitable were often not provided in the included studies. Regardless, authors mentioned multiple barriers and limitations that they had to account for, which differ per method and product. On the basis of the many lessons learned that we identified, there seems to be ample experience with, and knowledge about, different types of development methods. However, this knowledge remains mostly segregated and there could be more room for critical reflection on the suitability of a research activity. Furthermore, several potentially useful methods and products such as experience sampling, personas, or behavior change strategies seem to be underrepresented in the included studies. This underlines the need for better integration and broader dissemination of knowledge on eHealth development. Integrating and sharing knowledge can enable researchers to select the most fitting method or product, as opposed to using the one that is most easily available or well-known.

Comparison With Prior Work

Methods Versus Products

In this review, a distinction was made between methods and products in eHealth development. This distinction does justice to the diverging and converging nature of eHealth development. This is in line with design-thinking approaches such as the double diamond model, which pays attention to diverging, for example, by means of collecting data, and converging, for example, by means of integrating the findings in a set of requirements or a prototype [15,181]. Similarly, based on our findings, we recommend that methods and products should be seen not as separate activities but as 2 sides of the same eHealth development coin. This review showed that methods such as interviews, questionnaires, or literature studies are used to collect new data. These data can be translated into concrete products to synthesize collected data and to support and facilitate subsequent collection of new data. Thus, these products can serve as stepping stones among data collection methods. An example of using products as a synthesis approach is the formulation of values based on previously conducted interviews and questionnaires [15,97]. An example of the use of products to collect novel data is the use of a low-fidelity prototype in usability testing to gain insights into prospective users' needs and wishes [131]. The combination of methods and products illustrates how researchers can continuously check assumptions and decisions with stakeholders in a concrete, specific way by means of formative evaluation cycles [11]. However, currently, the terminology used in literature does not reflect the distinction among the different types of development activities. Every development-related activity is referred to as a method, or products are only briefly described or mentioned as a tool within a method. On the basis of this review, we suggest that researchers should make an explicit distinction between methods and products when reporting on their eHealth development process.

Iterative eHealth Development

This review showed that there is no single ideal way to conduct a development process. The included studies illustrate that there are multiple methods that can be used in many different ways, for many different goals, and at many different points in the development process. To illustrate, for identifying points of improvement in a current situation according to stakeholders, researchers might use interviews, focus groups, and questionnaires. In addition, values and requirements can be elicited by means of both an interview study and focus groups in which prototypes are evaluated [96,97,178,179]. Consequently, multiple methods can be used to reach similar goals. This implies that creating an ideal, step-by-step guideline with predetermined methods is neither possible nor desirable because this would suggest that there is a single most optimal way to conduct a development process. However, this conclusion does not mean that *anything goes*. On the basis of the lessons learned, it can be concluded that a specific method or product should fit the context in which the data collection takes place as well as the characteristics and skills of the participants and the outcomes of earlier development activities [182]. To illustrate, although in-depth interviews might be a useful way to gather information from health care workers and experts, they seem to be less suitable for people with dementia because of the cognitive skills that would be required [48,61,178]. In addition, although these approaches can be used to collect valuable data, they also illustrate that over the past 10 years, there has not been much innovation in the methods and products used in eHealth development. This is especially striking considering the rapid changes in the possibilities of eHealth technologies such as mobile apps or VR over the past years. Scientific methods such as literature reviews, focus groups, and interviews have been used very often, whereas innovative methods of a more participatory nature, for example, generative methods such as photo diaries and mood boards, are hardly used, despite their potential added value.

In addition, this review has shown that often, multiple methods can—and should—be combined to paint a full picture of the current situation and possibilities of eHealth. If done well, the use of multiple methods to incorporate the stakeholders' perspective throughout the entire process will result in an eHealth technology that fits their needs, wishes, skills, and context. This fit among technology, stakeholders, and context will also increase the chances of the uptake of eHealth in practice. Methods can be used in different ways to include the stakeholder perspective. On the one hand, methods can be used to validate findings; for example, interviews can be used to provide more context to the results of a literature review [15]. On the other hand, methods and products can also be used in a

complementary way; for example, although usability tests can be used to elicit requirements from the end user, they do not provide insight into the demands of the organization. This knowledge can be obtained by means of complementary focus groups with managers. Consequently, when selecting methods throughout the process, researchers should account for the suitability of the methods and products for their specific context, participants, and earlier findings of the development process. Furthermore, this review showed that not all methods used in human-centered development processes have to be participatory to be useful; for example, literature reviews or log data analyses can also be valuable. Nonparticipatory methods can be complemented with human-centered methods such as interviews or focus groups to form a coherent whole.

Finally, the lessons learned illustrated that when shaping a multi-method or mixed methods development process, an iterative and flexible development approach is key [33,54,55,64,113,115]. Ideally, decisions for the next development activities should be based on findings, new insights, and experiences that arose from previous activities. Such an iterative approach ensures that the development process fits the context and the people involved. This is in line with *agile science*, which states that eHealth development should be iterative and flexible, with short sprints to allow for constant changes to the process and products [14,183]. An agile approach requires a flexible mindset of researchers because it is often not possible to plan the entire development process from the start. To support researchers, the results of this review can serve as input for the creation of a *CeHRes Roadmap Toolkit* for selecting the most appropriate and fitting methods, as opposed to choosing the method that they are most familiar with. To broaden the scope of this toolkit, it could be complemented with other research, for example, reviews such as this one, but ones that are focused on other frameworks. Another way to broaden the toolkit is by asking experts on, for example, human-centered design and eHealth development about other potentially useful methods. Such a broader toolkit can support researchers in adapting their development process in case of contextual changes or unexpected or new findings [14,183]. This can facilitate the iterative nature of the development process.

Points of Improvement for eHealth Development Models

On the basis of the findings of this review, multiple points of improvement for the CeHRes Roadmap can be formulated. These can also be partly applied to other eHealth development models. First, it is important to note that most studies did not use all phases of the CeHRes Roadmap; rather, they *cherry-picked* the most relevant phases or principles to use in their work. Often, the CeHRes Roadmap was combined with other development approaches such as design thinking or the Behavior Change Wheel. This illustrates that in eHealth development, (parts of) multiple models, theories, and approaches can be used to complement each other [91]. Second, although business modeling and value specification are important elements of the CeHRes Roadmap, they were underrepresented in the included studies. This indicates that in general, researchers might require more guidelines on how to not only develop an intervention, but also think about its implementation in practice from the start of development. Third, a main point of improvement for eHealth development using the CeHRes Roadmap is related to the use of behavior change theory: this approach was underrepresented in the included studies [81]. This is quite surprising, especially considering the relationship between behavior change theory and effectiveness of (eHealth) interventions [184,185]. A possible explanation for this gap is that design-oriented, development models such as the CeHRes Roadmap do not explicitly *force* or nudge developers to incorporate theory, as opposed to more *human-centered* activities such as stakeholder identification or prototyping. Consequently, separate goals or activities might be added to the CeHRes Roadmap to support developers in incorporating behavior change theory [186,187]. In line with this, there is a need for more guidelines on how behavior change theory can be combined or complemented with persuasive design features. In addition, there is a need for more research on how this can in turn be connected to stakeholders' values and requirements [79,81,159]. An important point of attention in regard to this is that there is still much debate about the suitability of existing behavior change models for eHealth. Are existing theories suitable, or are they too static for highly personalized eHealth technologies? In other words: is there a need for more dynamic, personalized models for behavior change [188]? A final recommendation is related to applying development models to practice. As conducting development processes in practice is often very complex and challenging, multiple authors recommended forming a multidisciplinary development team. Such a team can consist of, for example, researchers, patients, health care providers, technology developers, designers, and managers [15,88,90]. A multidisciplinary team can prevent a top-down approach and tunnel vision; ensure that a constant eye is kept on the context; facilitate the coordination of large, complex development processes by combining skills and knowledge; and even support implementation of the to-be-developed eHealth technology in practice [182].

Strengths and Limitations

The main goal of this explorative narrative scoping review is to create an overview of development methods used in eHealth development guided by the CeHRes Roadmap. Although the CeHRes Roadmap is viewed as a suitable framework to guide eHealth development, there are many more useful development models. Because of our focus on studies that used the CeHRes Roadmap, several other suitable research methods or products might have been overlooked. Examples of such methods are observations, EMAs, or eye tracking [189-191]. Therefore, this overview of development methods and products is neither exhaustive nor unconditionally generalizable. Regardless, because of the interdisciplinary and broad applicability of the CeHRes Roadmap, much of the knowledge generated in this review is also relevant for other development models. However, creating a broader eHealth development toolkit would require additional similar reviews in which other development frameworks are put central. Subsequent analysis of differences and similarities among these reviews might also allow for generalization of knowledge. In line with this, because providing an elaborate, in-depth reflection on the included methods and products is beyond the scope of this paper, the provided lessons

learned in this narrative scoping review are also not exhaustive. To do justice to the possibilities of all mentioned methods and products, a separate review per method is required. In addition, because of the goal and broad nature of our scoping review, we did not assess the quality of the included studies. The main reason for this was that the quality of a study does not say anything about the quality of a development method or product. In other words, if a study in which interviews were used in a contextual inquiry is of extremely high methodological quality, this does not automatically mean that interviews in themselves are always the most suitable method for contextual inquiries. On the basis of the reported lessons learned, researchers have to make their own choice for the most suitable method or product for their process. Furthermore, a pitfall of the qualitative approach that was used to describe the methods and products is that it is prone to potential bias of the researchers. To overcome this, the screening and summarizing processes were set up in a systematic way and conducted by 3 researchers. In addition, all findings were discussed among, and critically reflected upon by, the coauthors—all with ample experience in multidisciplinary, human-centered eHealth development. Finally, as is the case in any review, not all relevant information on development methods might have been published. On the one hand, this might be because, too often, development studies are not viewed as full-fledged research and are simply not published. An example of this is a stakeholder analysis, which is often conducted but hardly ever included in publications [135]. On the other hand, we noticed that there is much difference between the clarity and extensiveness of the way authors reported and reflected on development activities. Suboptimal or incomplete reflection by authors resulted in some lessons learned that remained quite general and were not specifically focused on eHealth development. In general, this might have influenced our findings and highlights the need for a larger number of, and more comprehensive, papers on eHealth development.

Future Directions

As stated earlier, the results of this review should be viewed not as an exhaustive overview of all existing eHealth development methods and products but as an initial overview of activities related to research based on the CeHRes Roadmap. First, in multiple studies, the methods used were described in very general terms or the papers only described a single method from a larger development process [78]. Consequently, additional interviews, focus groups, or questionnaires—the choice of method naturally depending on the type of context and characteristics of the users—can be conducted with researchers to gain more in-depth insight into their experiences with eHealth development. This can also be extended to experts working in practice because innovation in human-centered design often is initiated in practice. In line with this, more specific reviews or viewpoint papers on single development methods or products might be conducted.

Second, a more general recommendation is related to the method of reporting. On the basis of the included studies, it becomes clear that there are many differences in the way authors report on development processes. In addition, as can be seen in the relatively low number of reported products compared with reported methods, there is room for improvement in the description of products used. We recommend that future development studies explicitly substantiate choices for, and critically reflect on, the methods and products used. Future research might explore the desirability and feasibility of a standardized guideline for development studies in line with the standards for reporting on, for example, randomized controlled trials (CONSORT [Consolidated Standards of Reporting Trials]) [192], reviews (PRISMA) [193], or qualitative research such as interviews (Consolidated Criteria for Reporting Qualitative Studies) [194]. On the basis of our findings, we recommend that researchers at least report on (1) the specific goal of the method, (2) a replicable description of the materials and procedure of data collection, (3) the rationale behind choosing the method, (4) the participants, (5) the setting, (6) the main results and conclusions, and (7) the most important lessons learned about the use of the method within the specific context [24]. In case of products, we recommend that they be described in a clear way and, if possible, that the products themselves be provided in the paper or appendix. Examples are screenshots of prototypes, a list with values, and an overview of stakeholders. In addition, multiple included studies showed that a visualization of the development process might also be of added value [15].

Third, the COVID-19 pandemic has shown us the potential of *online* research. Organizing web-based usability testing, focus groups, or interviews is sometimes not only necessary, but they can also have practical and even substantive advantages [87]. Future research can focus on if and how *research from a distance* can be applied to make eHealth development more efficient and effective. Finally, the results from this review can be used in the development of a digital *CeHRes Roadmap Toolkit*. This can be achieved by means of a human-centered development process in which end users—in this case, researchers and developers—are actively involved using suitable methods. Before a more comprehensive and exhaustive eHealth toolkit can be developed, there is a need for additional reviews that take a similar approach as this review but focus on other development models such as intervention mapping or user-centered design. This might result in a more exhaustive overview of other development methods and products that were not represented in this review. Furthermore, in line with the interdisciplinary nature of eHealth development, this eHealth toolkit can also be based on existing, more generic toolkits such as the Communication and Media Design methods pack [195] and the Human-Centered Design Toolkit [196]. Developing a broad eHealth toolkit requires an interdisciplinary approach in which designers with both a research and commercial background, psychologists, health scientists, other relevant stakeholders, and researchers combine their expertise. Including expertise from commercial design companies might be useful or even necessary to innovate eHealth development and go beyond the more traditional, scientific approaches.

Conclusions

This narrative scoping review has resulted in an initial overview of multiple methods and products that can be used for development of eHealth, guided by the CeHRes Roadmap. It can be concluded that there is not a single straightforward, step-by-step way to conduct a development process: the way a

process is shaped depends on the type of technology, context, and stakeholders. To support researchers in selecting the most appropriate and fitting methods and products for their development process, it is recommended to develop an eHealth development toolkit. To achieve this ultimate aim, there is a need not only for more research, but also for close collaboration with industry. In this way, eHealth development becomes less reliant on traditional research approaches such as literature reviews or interviews. Researchers can be challenged to create more creative and innovative development processes by using methods and products that are developed by industry. This review can serve as a blueprint for further reviews in which different frameworks are put central. In general, the decision for the method or product should be based on three different elements: (1) the specific goal of the current development phase, (2) the characteristics of the participants, and (3) practical requirements from the context. In this way, eHealth development processes can become more flexible and will be professionalized further. This in turn might result in more effective, personalized eHealth technologies that fit with, and are of added value for, the users and their context.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Data extraction form.
[DOCX File , 21 KB - jmir_v24i1e31858_app1.docx]

Multimedia Appendix 2
Main goal of the study and described methods and products per included record.
[PDF File (Adobe PDF File), 333 KB - jmir_v24i1e31858_app2.pdf]

References

1. Van Gemert-Pijnen L, Kip H, Kelders SM, Sanderman R. Introducing eHealth. In: eHealth Research, Theory and Development: A Multi-Disciplinary Approach. London: Routledge; 2018:3-26.
2. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec 05;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
3. Eysenbach G. What is e-health? J Med Internet Res 2001;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
4. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan 18;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
5. Van Gemert-Pijnen Kip L, Kelders SM, Kip H, Sanderman R, editors. eHealth Research, Theory and Development: A Multidisciplinary Approach. London: Routledge; 2018.
6. Kip H, Sieverink F, van Gemert-Pijnen JE, Bouman YH, Kelders SM. Integrating people, context, and technology in the implementation of a web-based intervention in forensic mental health care: mixed-methods study. J Med Internet Res 2020 May 26;22(5):e16906 [FREE Full text] [doi: 10.2196/16906] [Medline: 32348285]
7. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015 Jan 30;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
8. Beerlage-de Jong N, van Gemert-Pijnen L, Wentzel J, Hendrix R, Siemons L. Technology to support integrated antimicrobial stewardship programs: a user centered and stakeholder driven development approach. Infect Dis Rep 2017 Mar 30;9(1):6829 [FREE Full text] [doi: 10.4081/idr.2017.6829] [Medline: 28458796]
9. Eldredge L, Markham C, Ruiter R, Kok G, Parcel G. Planning Health Promotion Programs An Intervention Mapping Approach. New York: John Wiley & Sons; 2016.
10. Michie S, Yardley L, West R, Patrick K, Greaves F. Developing and evaluating digital interventions to promote behavior change in health and health care: recommendations resulting from an international workshop. J Med Internet Res 2017 Jun 29;19(6):e232 [FREE Full text] [doi: 10.2196/jmir.7126] [Medline: 28663162]
11. Kip H, Van Gemert-Pijnen L. Holistic development of ehealth technology. In: Van Gemert-Pijnen L, Kelders SM, Kip H, Sanderman R, editors. eHealth Research, Theory and Development: A Multi-Disciplinary Approach. London: Routledge; 2018:151-186.
12. Mohr DC, Lyon AR, Lattie EG, Reddy M, Schueller SM. Accelerating digital mental health research from early design and creation to successful implementation and sustainment. J Med Internet Res 2017 May 10;19(5):e153 [FREE Full text] [doi: 10.2196/jmir.7725] [Medline: 28490417]

13. Oinas-Kukkonen H, Harjumaa M. Persuasive systems design: key issues, process model, and system features. Commun Assoc Inf Syst 2009 Mar;24(1):28. [doi: 10.17705/1cais.02428]
14. Hekler EB, Klasnja P, Riley WT, Buman MP, Huberty J, Rivera DE, et al. Agile science: creating useful products for behavior change in the real world. Transl Behav Med 2016 Jun;6(2):317-328 [FREE Full text] [doi: 10.1007/s13142-016-0395-7] [Medline: 27357001]
15. Kip H, Kelders SM, Bouman YHA, van Gemert-Pijnen LJEWC. The Importance of Systematically Reporting and Reflecting on eHealth Development: Participatory Development Process of a Virtual Reality Application for Forensic Mental Health Care. J Med Internet Res 2019 Aug 19;21(8):e12972 [FREE Full text] [doi: 10.2196/12972] [Medline: 31429415]
16. Moore G, Wilding H, Gray K, Castle D. Participatory methods to engage health service users in the development of electronic health resources: systematic review. J Particip Med 2019 Feb 22;11(1):e11474 [FREE Full text] [doi: 10.2196/11474] [Medline: 33055069]
17. Maguire M. Methods to support human-centred design. Int J Human Comput Stud 2001 Oct;55(4):587-634. [doi: 10.1006/ijhc.2001.0503]
18. Phases of the CeHRes Roadmap. University of Twente. URL: https://www.utwente.nl/en/bms/ehealth/cehres-roadmap-toolkit/cehres-phases/ [accessed 2022-01-25]
19. Grant M, Booth A. A typology of reviews: an analysis of 14 review types and associated methodologies. Health Info Libr J 2009 Jun;26(2):91-108 [FREE Full text] [doi: 10.1111/j.1471-1842.2009.00848.x] [Medline: 19490148]
20. Van Dyk L, Wentzel M, Limburg, van M, Gemert-Pijnen, van JEWC, Schutte CS. Business models for sustained ehealth implementation: Lessons from two continents. In: Proceedings of International Conference on Computers and Industrial Engineering. 2012 Presented at: Proceedings of International Conference on Computers and Industrial Engineering; Jul 16-18, 2012; Cape Town, South Africa.
21. Van Gemert-Pijnen L, Span M. CeHRes roadmap to improve dementia care. In: Handbook of Smart Homes, Health Care and Well-Being. Cham: Springer; 2016.
22. Kooistra LC, Ruwaard J, Wiersma JE, van Oppen P, van der Vaart R, van Gemert-Pijnen JE, et al. Development and initial evaluation of blended cognitive behavioural treatment for major depression in routine specialized mental health care. Internet Interv 2016 May;4:61-71 [FREE Full text] [doi: 10.1016/j.invent.2016.01.003] [Medline: 30135791]
23. Kulyk O, Van Gemert-Pijnen L, Daas Den C, David S. How persuasive are serious games, social media and mHealth technologies for vulnerable young adults? Design factors for health behavior and lifestyle change support: sexual health case. In: Proceedings of the 3rd International Workshop on Behavior Change Support Systems, BCSS 2015 - co-located with the 10th International Conference on Persuasive Technology, PERSUASIVE 2015. 2015 Presented at: 3rd International Workshop on Behavior Change Support Systems, BCSS 2015 - co-located with the 10th International Conference on Persuasive Technology, PERSUASIVE 2015; Jun 3, 2015; Chicago, United States URL: https://abdn.pure.elsevier.com/en/publications/how-persuasive-are-serious-games-social-media-and-mhealth-technol
24. Ossebaard HC, Seydel ER, van Gemert-Pijnen J. Online usability and patients with long-term conditions: a mixed-methods approach. Int J Med Inform 2012 Jun;81(6):374-387. [doi: 10.1016/j.ijmedinf.2011.12.010] [Medline: 22261086]
25. Wentzel J, van Velsen L, van Limburg M, de Jong N, Karreman J, Hendrix R, et al. Participatory eHealth development to support nurses in antimicrobial stewardship. BMC Med Inform Decis Mak 2014 Jun 05;14:45 [FREE Full text] [doi: 10.1186/1472-6947-14-45] [Medline: 24898694]
26. Kulyk O, Akker RO, Klaassen R, Van Gemert-Pijnen J. Personalized virtual coaching for lifestyle support: principles for design and evaluation. Advances Life Sci 2014;6:300-309 [FREE Full text]
27. van Limburg M, Wentzel J, Sanderman R, van Gemert-Pijnen J. Business modeling to implement an eHealth portal for infection control: a reflection on co-creation with stakeholders. JMIR Res Protoc 2015 Aug 13;4(3):e104 [FREE Full text] [doi: 10.2196/resprot.4519] [Medline: 26272510]
28. Wentzel J, Müller F, Beerlage-de Jong N, van Gemert-Pijnen J. Card sorting to evaluate the robustness of the information architecture of a protocol website. Int J Med Inform 2016 Feb;86:71-81. [doi: 10.1016/j.ijmedinf.2015.12.003] [Medline: 26725698]
29. Beerlage-de Jong N, Eikelenboom-Boskamp A, Voss A, Sanderman R, van Gemert-Pijnen J. Combining user-centered design with the persuasive systems design model: the development process of a web-based registration and monitoring system for healthcare-associated infections in nursing homes. Int J Advances Life Sci 2014 Jan;6:262-271 [FREE Full text]
30. van Velsen L, Beaujean DJ, Wentzel J, Van Steenbergen JE, van Gemert-Pijnen JE. Developing requirements for a mobile app to support citizens in dealing with ticks and tick bites via end-user profiling. Health Informatics J 2015 Mar;21(1):24-35 [FREE Full text] [doi: 10.1177/1460458213496419] [Medline: 24282189]
31. Kelders SM, Pots WT, Oskam MJ, Bohlmeijer ET, van Gemert-Pijnen J. Development of a web-based intervention for the indicated prevention of depression. BMC Med Inform Decis Mak 2013 Feb 20;13:26 [FREE Full text] [doi: 10.1186/1472-6947-13-26] [Medline: 23425322]
32. Van de Dijk DD, Te Lintelo J, Willems CG, van Gemert-Pijnen J. ECOPD: user requirements of older people with copd for ehealth support at home, a user-centred study. In: Proceedings of the AAATE 2013. 2013 Presented at: AAATE 2013; Sep, 2013; Villamoura Portugal. [doi: 10.3233/978-1-61499-304-9-1279]

33. Gulmans J, Vollenbroek-Hutten M, Van Gemert-Pijnen J, Van Harten WH. Evaluating patient care communication in integrated care settings: application of a mixed method approach in cerebral palsy programs. Int J Qual Health Care 2009 Feb;21(1):58-65. [doi: 10.1093/intqhc/mzn053] [Medline: 19060037]
34. Verhoeven F, van Gemert-Pijnen J, Hendrix R. The development of a web-based tool for compliance with safe work practices; a case study about methicillin-resistant staphylococcus aureus. In: Proceedings of the 2009 International Conference on eHealth, Telemedicine, and Social Medicine. 2009 Presented at: 2009 International Conference on eHealth, Telemedicine, and Social Medicine; Feb 1-7, 2009; Cancun, Mexico. [doi: 10.1109/etelemed.2009.10]
35. Bossen D, Kloek C, Snippe HW, Dekker J, de Bakker D, Veenhof C. A blended intervention for patients with knee and hip osteoarthritis in the physical therapy practice: development and a pilot study. JMIR Res Protoc 2016 Feb 24;5(1):e32 [FREE Full text] [doi: 10.2196/resprot.5049] [Medline: 26912378]
36. Ehrler F, Lovis C, Blondon K. A mobile phone app for bedside nursing care: design and development using an adapted software development life cycle model. JMIR Mhealth Uhealth 2019 Apr 11;7(4):e12551 [FREE Full text] [doi: 10.2196/12551] [Medline: 30973339]
37. Cnossen IC, van Uden-Kraan CF, Eerenstein SE, Rinkel RN, Aalders IJ, van den Berg K, et al. A participatory design approach to develop a web-based self-care program supporting early rehabilitation among patients after total laryngectomy. Folia Phoniatr Logop 2015;67(4):193-201 [FREE Full text] [doi: 10.1159/000441251] [Medline: 26771305]
38. Harder H, Holroyd P, Burkinshaw L, Watten P, Zammit C, Harris PR, et al. A user-centred approach to developing bWell, a mobile app for arm and shoulder exercises after breast cancer treatment. J Cancer Surviv 2017 Dec;11(6):732-742 [FREE Full text] [doi: 10.1007/s11764-017-0630-3] [Medline: 28741202]
39. Honary M, Fisher NR, McNaney R, Lobban F. A web-based intervention for relatives of people experiencing psychosis or bipolar disorder: design study using a user-centered approach. JMIR Ment Health 2018 Dec 07;5(4):e11473 [FREE Full text] [doi: 10.2196/11473] [Medline: 30530457]
40. Ragouzeos D, Gandrup J, Berrean B, Li J, Murphy M, Trupin L, et al. "Am I OK?" using human centered design to empower rheumatoid arthritis patients through patient reported outcomes. Patient Educ Couns 2019 Mar;102(3):503-510 [FREE Full text] [doi: 10.1016/j.pec.2018.10.016] [Medline: 30446358]
41. Span M, Smits C, Jukema J, Groen-van de Ven L, Janssen R, Vernooij-Dassen M, et al. An interactive web tool for facilitating shared decision-making in dementia-care networks: a field study. Front Aging Neurosci 2015;7:128 [FREE Full text] [doi: 10.3389/fnagi.2015.00128] [Medline: 26217221]
42. Ammerlaan JW, Mulder OK, de Boer-Nijhof NC, Maat B, Kruize AA, van Laar J, et al. Building a tailored, patient-guided, web-based self-management intervention 'ReumaUitgedaagd!' for adults with a rheumatic disease: results of a usability study and design for a randomized control trail. JMIR Res Protoc 2016 Jun 23;5(2):e113 [FREE Full text] [doi: 10.2196/resprot.5735] [Medline: 27339472]
43. Corbett T, Walsh JC, Groarke A, Moss-Morris R, Morrissey E, McGuire BE. Cancer-related fatigue in post-treatment cancer survivors: theory-based development of a web-based intervention. JMIR Cancer 2017 Jul 04;3(2):e8 [FREE Full text] [doi: 10.2196/cancer.6987] [Medline: 28676465]
44. Limburg M, Ossebaard H, Wentzel M, van Gemert-Pijnen J. Co-creating with stakeholders: ideating eHealth applications to support antibiotic stewardship in hospitals. In: Proceedings of the eTELEMED 2016: The Eighth International Conference on eHealth, Telemedicine, and Social Medicine 2016. 2016 Presented at: eTELEMED 2016: The Eighth International Conference on eHealth, Telemedicine, and Social Medicine 2016; Apr 24 - 28, 2016; Venice, Italy.
45. Timmerman JG, Tönis TM, Dekker-van Weering MG, Stuiver MM, Wouters MW, van Harten WH, et al. Co-creation of an ICT-supported cancer rehabilitation application for resected lung cancer survivors: design and evaluation. BMC Health Serv Res 2016 Apr 27;16:155 [FREE Full text] [doi: 10.1186/s12913-016-1385-7] [Medline: 27121869]
46. McGrath SP, McGrath ML, Bastola D. Developing a concussion assessment mHealth app for certified Athletic Trainers. AMIA Annu Symp Proc 2017;2017:1282-1291 [FREE Full text] [Medline: 29854197]
47. Stuij SM, Labrie NH, van Dulmen S, Kersten MJ, Christoph N, Hulsman RL, INSTRUCT project group. Developing a digital communication training tool on information-provision in oncology: uncovering learning needs and training preferences. BMC Med Educ 2018 Sep 24;18(1):220 [FREE Full text] [doi: 10.1186/s12909-018-1308-x] [Medline: 30249221]
48. Span M, Hettinga M, Smits C, Groen-van de Ven L, Jukema J, Vernooij-Dassen M, et al. Developing a supportive tool to facilitate shared decision making in dementia. In: Proceedings of the The Sixth International Conference on eHealth, Telemedicine, and Social Medicine. 2014 Presented at: The Sixth International Conference on eHealth, Telemedicine, and Social Medicine; Mar 23-27, 2014; Barcelona, Spain.
49. Pearson J, Walsh N, Carter D, Koskela S, Hurley M. Developing a web-based version of an exercise-based rehabilitation program for people with chronic knee and hip pain: a mixed methods study. JMIR Res Protoc 2016 May 19;5(2):e67 [FREE Full text] [doi: 10.2196/resprot.5446] [Medline: 27197702]
50. Fledderus M, Schreurs KM, Bohlmeijer ET, Vollenbroek-Hutten MM. Development and pilot evaluation of an online relapse-prevention program based on acceptance and commitment therapy for chronic pain patients. JMIR Hum Factors 2015 Jan 05;2(1):e1 [FREE Full text] [doi: 10.2196/humanfactors.3302] [Medline: 27025550]

51. Kloek C, van Tilburg M, Staal J, Veenhof C, Bossen D. Development and proof of concept of a blended physiotherapeutic intervention for patients with non-specific low back pain. Physiotherapy 2019 Dec;105(4):483-491 [FREE Full text] [doi: 10.1016/j.physio.2018.12.006] [Medline: 31031023]
52. Wachtler C, Coe A, Davidson S, Fletcher S, Mendoza A, Sterling L, et al. Development of a Mobile Clinical Prediction Tool to Estimate Future Depression Severity and Guide Treatment in Primary Care: User-Centered Design. JMIR Mhealth Uhealth 2018 Apr 23;6(4):e95. [doi: 10.2196/mhealth.9502]
53. Maguire R, Ream E, Richardson A, Connaghan J, Johnston B, Kotronoulas G, et al. Development of a novel remote patient monitoring system: the advanced symptom management system for radiotherapy to improve the symptom experience of patients with lung cancer receiving radiotherapy. Cancer Nurs 2015;38(2):E37-E47. [doi: 10.1097/NCC.0000000000000150] [Medline: 24836956]
54. Al-Itejawi HH, van Uden-Kraan CF, Vis AN, Nieuwenhuijzen JA, Hofstee MJ, van Moorselaar RJ, et al. Development of a patient decision aid for the treatment of localised prostate cancer: a participatory design approach. J Clin Nurs 2016 Apr;25(7-8):1131-1144. [doi: 10.1111/jocn.13120] [Medline: 26880526]
55. Reblin M, Wu Y, Pok J, Kane L, Colman H, Cohen AL, et al. Development of the electronic social network assessment program using the center for eHealth and wellbeing research roadmap. JMIR Hum Factors 2017 Aug 30;4(3):e23 [FREE Full text] [doi: 10.2196/humanfactors.7845] [Medline: 28855149]
56. Gilbert M, Haag D, Hottes TS, Bondyra M, Elliot E, Chabot C, et al. Get checked… where? The development of a comprehensive, integrated internet-based testing program for sexually transmitted and blood-borne infections in British Columbia, Canada. JMIR Res Protoc 2016 Sep 20;5(3):e186 [FREE Full text] [doi: 10.2196/resprot.6293] [Medline: 27649716]
57. Wentink M, van Bodegom-Vos L, Brouns B, Arwert H, Houdijk S, Kewalbansing P, et al. How to improve eRehabilitation programs in stroke care? A focus group study to identify requirements of end-users. BMC Med Inform Decis Mak 2019 Jul 26;19(1):145 [FREE Full text] [doi: 10.1186/s12911-019-0871-3] [Medline: 31349824]
58. Skjøth MM, Hansen HP, Draborg E, Pedersen CD, Lamont RF, Jørgensen JS. Informed choice for participation in down syndrome screening: development and content of a web-based decision aid. JMIR Res Protoc 2015 Sep 21;4(3):e113 [FREE Full text] [doi: 10.2196/resprot.4291] [Medline: 26392319]
59. Askedal K, Flak L. Stakeholder contradictions in early stages of eHealth efforts. In: Proceedings of the 50th Hawaii International Conference on System Sciences. 2017 Presented at: 50th Hawaii International Conference on System Sciences; 2017; Hawaii. [doi: 10.24251/hicss.2017.415]
60. Galpin A, Meredith J, Ure C, Robinson L. "Thanks for letting us all share your mammogram experience virtually": developing a web-based hub for breast cancer screening. JMIR Cancer 2017 Oct 27;3(2):e17 [FREE Full text] [doi: 10.2196/cancer.8150] [Medline: 29079555]
61. Span M, Smits C, Groen-van de Ven L, Jukema J, Cremers A, Vernooij-Dassen M, et al. Towards an interactive web tool that supports shared decision making in dementia: identifying user requirements. Int J Adv Life Sci 2014;6(3&4):338-349.
62. Fico G, Teresa AM. Use of an holistic approach for effective adoption of User-Centred-Design techniques in diabetes disease management: experiences in user need elicitation. In: Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2015 Presented at: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); Aug 25-29, 2015; Milan, Italy. [doi: 10.1109/embc.2015.7318812]
63. Fico G, Hernandez L, Cancela J, Arredondo M. User Centered Design to incorporate predictive models for Type 2 Diabetes screening and management into professional decision support tools: preliminary results. In: Proceedings of the World Congress on Medical Physics and Biomedical Engineering. 2015 Presented at: World Congress on Medical Physics and Biomedical Engineering; Jun 7-12, 2015; Toronto, Canada p. 1389-1392. [doi: 10.1007/978-3-319-19387-8_338]
64. Fico G, Hernanzez L, Cancela J, Dagliati A, Sacchi L, Martinez-Millana A, et al. What do healthcare professionals need to turn risk models for type 2 diabetes into usable computerized clinical decision support systems? Lessons learned from the MOSAIC project. BMC Med Inform Decis Mak 2019 Aug 16;19(1):163 [FREE Full text] [doi: 10.1186/s12911-019-0887-8] [Medline: 31419982]
65. Sjölinder M, Ehn M, Boman I, Folke M, Hansson P, Sommerfeld D, et al. A multi-disciplinary approach in the development of a stroke rehabilitation tool. In: Human-Computer Interaction. Cham: Springer; 2014.
66. Ledel Solem IK, Varsi C, Eide H, Kristjansdottir OB, Børøsund E, Schreurs KM, et al. A user-centered approach to an evidence-based electronic health pain management intervention for people with chronic pain: design and development of EPIO. J Med Internet Res 2020 Jan 21;22(1):e15889 [FREE Full text] [doi: 10.2196/15889] [Medline: 31961331]
67. Imhemed S, Kamaruddin A, Hasan S. Designing a Mobile App to support a Healthy Lifestyle (MAS-HeaL) for type-2 diabetes patients in Malaysia. J Eng Applied Sci 2017;12(3):559-563.
68. Tieman JJ, Morgan DD, Swetenham K, To TH, Currow DC. Designing clinically valuable telehealth resources: processes to develop a community-based palliative care prototype. JMIR Res Protoc 2014 Sep 04;3(3):e41 [FREE Full text] [doi: 10.2196/resprot.3266] [Medline: 25189279]

69. Mummah SA, King AC, Gardner CD, Sutton S. Iterative development of Vegethon: a theory-based mobile app intervention to increase vegetable consumption. Int J Behav Nutr Phys Act 2016 Aug 08;13:90 [FREE Full text] [doi: 10.1186/s12966-016-0400-z] [Medline: 27501724]
70. Derks YP, Visser TD, Bohlmeijer ET, Noordzij ML. mHealth in Mental Health: how to efficiently and scientifically create an ambulatory biofeedback e-coaching app for patients with borderline personality disorder. Int J Human Factor Ergonom 2017;5(1):61-92. [doi: 10.1504/ijhfe.2017.088418]
71. Cheung KL, Hiligsmann M, Präger M, Jones T, Józwiak-Hagymásy J, Muñoz C, et al. Optimizing usability of an economic decision support tool: prototype of the equipt tool. Int J Technol Assess Health Care 2018 Jan;34(1):68-77. [doi: 10.1017/S0266462317004470] [Medline: 29455684]
72. Yap MB, Lawrence KA, Rapee RM, Cardamone-Breen MC, Green J, Jorm AF. Partners in parenting: a multi-level web-based approach to support parents in prevention and early intervention for adolescent depression and anxiety. JMIR Ment Health 2017 Dec 19;4(4):e59 [FREE Full text] [doi: 10.2196/mental.8492] [Medline: 29258974]
73. Dick S, Heavin C, Vasiliou V. The design of a digital behaviour change intervention for third-level student illicit substance use: a persona building approach. In: Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020 Presented at: 53rd Hawaii International Conference on System Sciences; Jan 7-10, 2020; Hawaii. [doi: 10.24251/hicss.2020.415]
74. De Backere F, Ongenae F, Vannieuwenborg F, Van Ooteghem J, Duysburgh P, Jansen A, et al. The OCareCloudS project: toward organizing care through trusted cloud services. Inform Health Soc Care 2016;41(2):159-176. [doi: 10.3109/17538157.2014.965306] [Medline: 25325572]
75. Vermeulen J, Neyens J, Spreeuwenberg MD, van Rossum E, Sipers W, Habets H, et al. User-centered development and testing of a monitoring system that provides feedback regarding physical functioning to elderly people. Patient Prefer Adherence 2013;7:843-854 [FREE Full text] [doi: 10.2147/PPA.S45897] [Medline: 24039407]
76. Kuosmanen T, Fleming TM, Barry MM. Using computerized mental health programs in alternative education: understanding the requirements of students and staff. Health Commun 2018 Jun;33(6):753-761. [doi: 10.1080/10410236.2017.1309620] [Medline: 28426253]
77. Verhoeven F, van Gemert-Pijnen J. Discount user-centered e-health design: a quick-but-not-dirty method. In: HCI in Work and Learning, Life and Leisure. Berlin, Heidelberg: Springer; 2010.
78. Amann J, Fiordelli M, Scheel-Sailer A, Brach M, Rubinelli S. Opportunities and challenges of a self-management app to support people with spinal cord injury in the prevention of pressure injuries: qualitative study. JMIR Mhealth Uhealth 2020 Dec 09;8(12):e22452 [FREE Full text] [doi: 10.2196/22452] [Medline: 33295876]
79. Asbjørnsen RA, Wentzel J, Smedsrød ML, Hjelmesæth J, Clark MM, Solberg Nes L, et al. Identifying persuasive design principles and behavior change techniques supporting end user values and needs in eHealth interventions for long-term weight loss maintenance: qualitative study. J Med Internet Res 2020 Nov 30;22(11):e22598 [FREE Full text] [doi: 10.2196/22598] [Medline: 33252347]
80. Bogza L, Patry-Lebeau C, Farmanova E, Witteman HO, Elliott J, Stolee P, et al. User-centered design and evaluation of a web-based decision aid for older adults living with mild cognitive impairment and their health care providers: mixed methods study. J Med Internet Res 2020 Aug 19;22(8):e17406 [FREE Full text] [doi: 10.2196/17406] [Medline: 32442151]
81. Breeman LD, Keesman M, Atsma DE, Chavannes NH, Janssen V, van Gemert-Pijnen L, BENEFIT consortium. A multi-stakeholder approach to eHealth development: promoting sustained healthy living among cardiovascular patients. Int J Med Inform 2021 Mar;147:104364 [FREE Full text] [doi: 10.1016/j.ijmedinf.2020.104364] [Medline: 33373949]
82. D'Addario M, Baretta D, Zanatta F, Greco A, Steca P. Engagement features in physical activity smartphone apps: focus group study with sedentary people. JMIR Mhealth Uhealth 2020 Nov 16;8(11):e20460 [FREE Full text] [doi: 10.2196/20460] [Medline: 33196450]
83. Jonathan GK, Dopke CA, Michaels T, Bank A, Martin CR, Adhikari K, et al. A smartphone-based self-management intervention for bipolar disorder (LiveWell): user-centered development approach. JMIR Ment Health 2021 Apr 12;8(4):e20424 [FREE Full text] [doi: 10.2196/20424] [Medline: 33843607]
84. Korpershoek YJ, Hermsen S, Schoonhoven L, Schuurmans MJ, Trappenburg JC. User-centered design of a mobile health intervention to enhance exacerbation-related self-management in patients with chronic obstructive pulmonary disease (Copilot): mixed methods study. J Med Internet Res 2020 Jun 15;22(6):e15449 [FREE Full text] [doi: 10.2196/15449] [Medline: 32538793]
85. Liu LH, Garrett SB, Li J, Ragouzeos D, Berrean B, Dohan D, et al. Patient and clinician perspectives on a patient-facing dashboard that visualizes patient reported outcomes in rheumatoid arthritis. Health Expect 2020 Aug;23(4):846-859 [FREE Full text] [doi: 10.1111/hex.13057] [Medline: 32270591]
86. Nicklas JM, Leiferman JA, Lockhart S, Daly KM, Bull SS, Barbour LA. Development and modification of a mobile health program to promote postpartum weight loss in women at elevated risk for cardiometabolic disease: single-arm pilot study. JMIR Form Res 2020 Apr 09;4(4):e16151 [FREE Full text] [doi: 10.2196/16151] [Medline: 32271149]
87. Oberschmidt K, Grünloh C, Tunç S, Van VL, Nijboer F. You can't always get what you want: streamlining stakeholder interests when designing technology-supported services for Active and Assisted Living. In: Proceedings of the 32nd

Australian Conference on Human-Computer Interaction. 2020 Presented at: 32nd Australian Conference on Human-Computer Interaction; Dec 2 - 4, 2020; Sydney NSW Australia. [doi: 10.1145/3441000.3441040]
88. Rai HK, Schneider J, Orrell M. An individual cognitive stimulation therapy app for people with dementia: development and usability study of thinkability. JMIR Aging 2020 Nov 16;3(2):e17105 [FREE Full text] [doi: 10.2196/17105] [Medline: 33196451]
89. Rai HK, Griffiths R, Yates L, Schneider J, Orrell M. Field-testing an iCST touch-screen application with people with dementia and carers: a mixed method study. Aging Ment Health 2021 Jun;25(6):1008-1018. [doi: 10.1080/13607863.2020.1783515] [Medline: 32578445]
90. Schmidt T, Kok R, Andersen CM, Skovbakke SJ, Ahm R, Wiil UK, et al. Development of an internet-delivered program and platform for the treatment of depression and anxiety in patients with ischemic heart disease in eMindYourHeart. Inform Health Soc Care 2021 Jun 02;46(2):178-191. [doi: 10.1080/17538157.2021.1878185] [Medline: 33632054]
91. Sporrel K, De Boer RD, Wang S, Nibbeling N, Simons M, Deutekom M, et al. The design and development of a personalized leisure time physical activity application based on behavior change theories, end-user perceptions, and principles from empirical data mining. Front Public Health 2020;8:528472 [FREE Full text] [doi: 10.3389/fpubh.2020.528472] [Medline: 33604321]
92. Stuij SM, Drossaert CH, Labrie NH, Hulsman RL, Kersten MJ, van Dulmen S, INSTRUCT project group. Developing a digital training tool to support oncologists in the skill of information-provision: a user centred approach. BMC Med Educ 2020 May 01;20(1):135 [FREE Full text] [doi: 10.1186/s12909-020-1985-0] [Medline: 32357886]
93. Timmer MA, Kloek CJ, de Kleijn P, Kuijlaars IA, Schutgens RE, Veenhof C, et al. A blended physiotherapy intervention for persons with hemophilic arthropathy: development study. J Med Internet Res 2020 Jun 19;22(6):e16631 [FREE Full text] [doi: 10.2196/16631] [Medline: 32558654]
94. Winterling J, Wiklander M, Obol CM, Lampic C, Eriksson LE, Pelters B, et al. Development of a self-help web-based intervention targeting young cancer patients with sexual problems and fertility distress in collaboration with patient research partners. JMIR Res Protoc 2016 Apr 12;5(2):e60 [FREE Full text] [doi: 10.2196/resprot.5499] [Medline: 27073057]
95. Carr EC, Babione JN, Marshall D. Translating research into practice through user-centered design: an application for osteoarthritis healthcare planning. Int J Med Inform 2017 Aug;104:31-37. [doi: 10.1016/j.ijmedinf.2017.05.007] [Medline: 28599814]
96. Kip H, Kelders SM, Weerink K, Kuiper A, Brüninghoff I, Bouman YH, et al. Identifying the added value of virtual reality for treatment in forensic mental health: a scenario-based, qualitative approach. Front Psychol 2019;10:406 [FREE Full text] [doi: 10.3389/fpsyg.2019.00406] [Medline: 30873093]
97. Kip H, Kelders S, Van Gemert-Pijnen J. Putting the value in VR: how to systematically and iteratively develop a value-based vr application with a complex target group. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK. [doi: 10.1145/3290605.3300365]
98. Verhoeven F, Hendrix RM, Daniels-Haardt I, Friedrich AW, Van Gemert-Pijnen JE. The development of a web-based learning tool for prevention and control of MRSA. Hygiene Medizin 2008 Oct;33(10):412-416 [FREE Full text]
99. van Limburg AH, Van Gemert-Pijnen J. Towards innovative business modeling for sustainable eHealth applications. In: Proceedings of the 2010 Second International Conference on eHealth, Telemedicine, and Social Medicine. 2010 Presented at: 2010 Second International Conference on eHealth, Telemedicine, and Social Medicine; Feb 10-16, 2010; Saint Maarten, Netherlands Antilles. [doi: 10.1109/etelemed.2010.30]
100. Beaujean D, van Velsen L, van Gemert-Pijnen JE, Maat A, van Steenbergen JE, Crutzen R. Using risk group profiles as a lightweight qualitative approach for intervention development: an example of prevention of tick bites and lyme disease. JMIR Res Protoc 2013 Oct 30;2(2):e45 [FREE Full text] [doi: 10.2196/resprot.2760] [Medline: 24172875]
101. van der Vaart R, Witting M, Riper H, Kooistra L, Bohlmeijer ET, van Gemert-Pijnen LJ. Blending online therapy into regular face-to-face therapy for depression: content, ratio and preconditions according to patients and therapists using a Delphi study. BMC Psychiatry 2014 Dec 14;14(1):355-310 [FREE Full text] [doi: 10.1186/s12888-014-0355-z] [Medline: 25496393]
102. van Gemert-Pijnen J, Karreman J, Vonderhorst S, Verhoeven F, Wentzel J. Participatory development via user-involvement - A case study about the development of a web-based patient-communication system about methicillin-resistant staphylococcus aureus. Electron J Health Informatics 2011;6(4):e28. [doi: 10.1109/etelemed.2010.31]
103. Lentferink A, Polstra L, de Groot M, Oldenhuis H, Velthuijsen H, van Gemert-Pijnen J. The values of self-tracking and persuasive eCoaching according to employees and human resource advisors for a workplace stress management application: a qualitative study. In: Persuasive Technology. Cham: Springer; 2018.
104. Cruz-Martínez R, Sieverink F, Wesselink R, van Gemert-Pijnen J. Towards data-driven persuasive coaching in a heart failure telemonitoring technology. In: Proceedings of the Sixth International Workshop on Behavior Change Support Systems (BCSS 2018). In conjunction with the 13th International Conference on Persuasive Technology. 2018 Presented at: Sixth International Workshop on Behavior Change Support Systems (BCSS 2018). In conjunction with the 13th International Conference on Persuasive Technology; Apr 16-19, 2018; University of Waterloo, Canada.

105. Teles S, Bertal D, Kofler A, Ruscher S, Paul C. A multi-perspective view on AAL stakeholders' needs a user-centred requirement analysis for the activeadvice European project. In: Proceedings of the 3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health - ICT4AWE. 2017 Presented at: Proceedings of the 3rd International Conference on Information and Communication Technologies for Ageing Well and e-Health - ICT4AWE; Apr 28-29, 2017; Porto, Portugal. [doi: 10.5220/0006380701040116]
106. Børøsund E, Mirkovic J, Clark MM, Ehlers SL, Andrykowski MA, Bergland A, et al. A stress management app intervention for cancer survivors: design, development, and usability testing. JMIR Form Res 2018 Sep 06;2(2):e19 [FREE Full text] [doi: 10.2196/formative.9954] [Medline: 30684438]
107. Nobakht Z, Rassafiani M, Hosseini SA. A web-based caring training for caregivers of children with cerebral palsy: development and evaluation. Iran J Child Neurol 2018;12(4):65-84 [FREE Full text] [Medline: 30279710]
108. Suffoletto B, Yanta J, Kurtz R, Cochran G, Douaihy A, Chung T. Acceptability of an opioid relapse prevention text-message intervention for emergency department patients. J Addict Med 2017;11(6):475-482 [FREE Full text] [doi: 10.1097/ADM.0000000000000351] [Medline: 28858888]
109. van Bruinessen IR, van Weel-Baumgarten EM, Snippe HW, Gouw H, Zijlstra JM, van Dulmen S. Active patient participation in the development of an online intervention. JMIR Res Protoc 2014 Nov 06;3(4):e59 [FREE Full text] [doi: 10.2196/resprot.3695] [Medline: 25379679]
110. Duman-Lubberding S, van Uden-Kraan CF, Peek N, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. An eHealth application in head and neck cancer survivorship care: health care professionals' perspectives. J Med Internet Res 2015 Oct 21;17(10):e235 [FREE Full text] [doi: 10.2196/jmir.4870] [Medline: 26489918]
111. Span M, Smits C, Groen-Van de Ven LM, Cremers A, Jukema J, Vernooij-Dassen M, et al. An interactive web tool to facilitate shared decision making in dementia: design issues perceived by caregivers and patients. Int J Adv Life Sci 2014;6(3-4):107-121 [FREE Full text]
112. Hochstenbach LM, Courtens AM, Zwakhalen SM, Vermeulen J, van Kleef M, de Witte LP. Co-creative development of an eHealth nursing intervention: self-management support for outpatients with cancer pain. Appl Nurs Res 2017 Aug;36:1-8. [doi: 10.1016/j.apnr.2017.03.004] [Medline: 28720227]
113. Rothgangel A, Braun S, Smeets R, Beurskens A. Design and development of a telerehabilitation platform for patients with phantom limb pain: a user-centered approach. JMIR Rehabil Assist Technol 2017 Feb 15;4(1):e2 [FREE Full text] [doi: 10.2196/rehab.6761] [Medline: 28582249]
114. Walters LE, Scott RE, Mars M. Design requirements for a teledermatology scale-up framework. South African Comput J 2018 Jul 10;30(1):128-160. [doi: 10.18489/sacj.v30i1.559]
115. Doty JL, Brady SS, Monardez Popelka J, Rietveld L, Garcia-Huidobro D, Doty MJ, et al. Designing a mobile app to enhance parenting skills of Latinx parents: a community-based participatory approach. JMIR Form Res 2020 Jan 24;4(1):e12618 [FREE Full text] [doi: 10.2196/12618] [Medline: 32012034]
116. Pang PC, Chang S, Verspoor K, Pearce J. Designing health websites based on users' web-based information-seeking behaviors: a mixed-method observational study. J Med Internet Res 2016 Jun 06;18(6):e145 [FREE Full text] [doi: 10.2196/jmir.5661] [Medline: 27267955]
117. Geraghty AW, Muñoz RF, Yardley L, Mc Sharry J, Little P, Moore M. Developing an unguided internet-delivered intervention for emotional distress in primary care patients: applying common factor and person-based approaches. JMIR Ment Health 2016 Dec 20;3(4):e53 [FREE Full text] [doi: 10.2196/mental.5845] [Medline: 27998878]
118. Crane D, Garnett C, Brown J, West R, Michie S. Factors influencing usability of a smartphone app to reduce excessive alcohol consumption: think aloud and interview studies. Front Public Health 2017;5:39 [FREE Full text] [doi: 10.3389/fpubh.2017.00039] [Medline: 28421175]
119. Lubberding S, van Uden-Kraan CF, Te Velde EA, Cuijpers P, Leemans CR, Verdonck-de Leeuw IM. Improving access to supportive cancer care through an eHealth application: a qualitative needs assessment among cancer survivors. J Clin Nurs 2015 May;24(9-10):1367-1379. [doi: 10.1111/jocn.12753] [Medline: 25677218]
120. Downing KL, Best K, Campbell KJ, Hesketh KD. Informing active play and screen time behaviour change interventions for low socioeconomic position mothers of young children: what do mothers want? Biomed Res Int 2016;2016:2139782 [FREE Full text] [doi: 10.1155/2016/2139782] [Medline: 28053979]
121. Noordman J, Driesenaar JA, van Bruinessen IR, van Dulmen S. ListeningTime; participatory development of a web-based preparatory communication tool for elderly cancer patients and their healthcare providers. Internet Interv 2017 Sep;9:51-56 [FREE Full text] [doi: 10.1016/j.invent.2017.05.002] [Medline: 30135837]
122. Boele FW, van Uden-Kraan CF, Hilverda K, Weimer J, Donovan HS, Drappatz J, et al. Neuro-oncology family caregivers' view on keeping track of care issues using eHealth systems: it's a question of time. J Neurooncol 2017 Aug;134(1):157-167 [FREE Full text] [doi: 10.1007/s11060-017-2504-y] [Medline: 28550505]
123. Merolli M, Marshall CJ, Pranata A, Paay J, Sterling L. User-centered value specifications for technologies supporting chronic low-back pain management. Stud Health Technol Inform 2019 Aug 21;264:1288-1292. [doi: 10.3233/SHTI190434] [Medline: 31438133]

124. Floch J, Zettl A, Fricke L, Weisser T, Grut L, Vilarinho T, et al. User needs in the development of a health app ecosystem for self-management of cystic fibrosis: user-centered development approach. JMIR Mhealth Uhealth 2018 May 08;6(5):e113 [FREE Full text] [doi: 10.2196/mhealth.8236] [Medline: 29739742]
125. van der Horst H. Vreemde vogels. HUWE 2005 Mar;48(3):883. [doi: 10.1007/bf03084505]
126. Liao G, Chien Y, Chen Y, Hsiung H, Chen H, Hsieh M, et al. What to build for middle-agers to come? Attractive and necessary functions of exercise-promotion mobile phone apps: a cross-sectional study. JMIR Mhealth Uhealth 2017 May 25;5(5):e65 [FREE Full text] [doi: 10.2196/mhealth.6233] [Medline: 28546140]
127. Derks YP, Klaassen R, Westerhof GJ, Bohlmeijer ET, Noordzij ML. Development of an ambulatory biofeedback app to enhance emotional awareness in patients with borderline personality disorder: multicycle usability testing study. JMIR Mhealth Uhealth 2019 Oct 15;7(10):e13479 [FREE Full text] [doi: 10.2196/13479] [Medline: 31617851]
128. Keizer J, Beerlage-De Jong N, Al Naiemi N, van Gemert-Pijnen J. Finding the match between healthcare worker and expert for optimal audit and feedback on antimicrobial resistance prevention measures. Antimicrob Resist Infect Control 2020 Aug 05;9(1):125 [FREE Full text] [doi: 10.1186/s13756-020-00794-7] [Medline: 32758300]
129. Lentferink A, Polstra L, D'Souza A, Oldenhuis H, Velthuijsen H, van Gemert-Pijnen L. Creating value with eHealth: identification of the value proposition with key stakeholders for the resilience navigator app. BMC Med Inform Decis Mak 2020 Apr 27;20(1):76 [FREE Full text] [doi: 10.1186/s12911-020-1088-1] [Medline: 32340615]
130. van Velsen L, van Gemert-Pijnen J, Beaujean DJ, Wentzel J, van Steenbergen JE. Should health organizations use web 2.0 media in times of an infectious disease crisis? An in-depth qualitative study of citizens' information behavior during an EHEC outbreak. J Med Internet Res 2012 Dec 20;14(6):e181 [FREE Full text] [doi: 10.2196/jmir.2123] [Medline: 23257066]
131. Beerlage-de Jong N, Wentzel J, Hendrix R, van Gemert-Pijnen L. The value of participatory development to support antimicrobial stewardship with a clinical decision support system. Am J Infect Control 2017 Apr 01;45(4):365-371. [doi: 10.1016/j.ajic.2016.12.001] [Medline: 28089673]
132. Scholten MR, Kelders SM, Van Gemert-Pijnen J. An empirical study of a pedagogical agent as an adjunct to an eHealth self-management intervention: what modalities does it need to successfully support and motivate users? Front Psychol 2019;10:1063 [FREE Full text] [doi: 10.3389/fpsyg.2019.01063] [Medline: 31143152]
133. Nijland N, Cranen K, Boer H, van Gemert-Pijnen J, Seydel ER. Patient use and compliance with medical advice delivered by a web-based triage system in primary care. J Telemed Telecare 2010;16(1):8-11. [doi: 10.1258/jtt.2009.001004] [Medline: 20086260]
134. van Velsen L, Beaujean DJ, van Gemert-Pijnen JE, van Steenbergen JE, Timen A. Public knowledge and preventive behavior during a large-scale Salmonella outbreak: results from an online survey in the Netherlands. BMC Public Health 2014 Jan 31;14:100 [FREE Full text] [doi: 10.1186/1471-2458-14-100] [Medline: 24479614]
135. van Woezik AF, Braakman-Jansen LM, Kulyk O, Siemons L, van Gemert-Pijnen JE. Tackling wicked problems in infection prevention and control: a guideline for co-creation with stakeholders. Antimicrob Resist Infect Control 2016;5:20 [FREE Full text] [doi: 10.1186/s13756-016-0119-2] [Medline: 27213040]
136. Bernecker SL, Banschback K, Santorelli GD, Constantino MJ. A web-disseminated self-help and peer support program could fill gaps in mental health care: lessons from a consumer survey. JMIR Ment Health 2017 Jan 19;4(1):e5 [FREE Full text] [doi: 10.2196/mental.4751] [Medline: 28104578]
137. Mirk SM, Wegrzyn NM. Apps for health-related education in pharmacy practice: needs assessment survey among patients within a large metropolitan area. JMIR Res Protoc 2017 Jul 19;6(7):e133 [FREE Full text] [doi: 10.2196/resprot.5886] [Medline: 28724511]
138. Jansen F, van Uden-Kraan CF, van Zwieten V, Witte BI, Verdonck-de Leeuw IM. Cancer survivors' perceived need for supportive care and their attitude towards self-management and eHealth. Support Care Cancer 2015 Jun 26;23(6):1679-1688. [doi: 10.1007/s00520-014-2514-7] [Medline: 25424520]
139. Whiteside U, Lungu A, Richards J, Simon GE, Clingan S, Siler J, et al. Designing messaging to engage patients in an online suicide prevention intervention: survey results from patients with current suicidal ideation. J Med Internet Res 2014 Feb 07;16(2):e42 [FREE Full text] [doi: 10.2196/jmir.3173] [Medline: 24509475]
140. Noordman J, Driesenaar JA, van Bruinessen IR, Portielje JE, van Dulmen S. Evaluation and implementation of ListeningTime: a web-based preparatory communication tool for elderly patients with cancer and their health care providers. JMIR Cancer 2019 Jan 30;5(1):e11556 [FREE Full text] [doi: 10.2196/11556] [Medline: 30698525]
141. van Uden-Kraan CF, Jansen F, Lissenberg-Witte BI, Eerenstein SE, Leemans CR, Verdonck-de Leeuw IM. Health-related and cancer-related Internet use by patients treated with total laryngectomy. Support Care Cancer 2020 Jan;28(1):131-140 [FREE Full text] [doi: 10.1007/s00520-019-04757-6] [Medline: 30993449]
142. Da Costa D, Zelkowitz P, Letourneau N, Howlett A, Dennis C, Russell B, et al. HealthyDads.ca: what do men want in a website designed to promote emotional wellness and healthy behaviors during the transition to parenthood? J Med Internet Res 2017 Oct 11;19(10):e325 [FREE Full text] [doi: 10.2196/jmir.7415] [Medline: 29021126]
143. Grunberg PH, Dennis C, Da Costa D, Zelkowitz P. Infertility patients' need and preferences for online peer support. Reprod Biomed Soc Online 2018 Aug;6:80-89 [FREE Full text] [doi: 10.1016/j.rbms.2018.10.016] [Medline: 30547107]

144. Van de Belt TH, Engelen LJ, Berben SA, Teerenstra S, Samsom M, Schoonhoven L. Internet and social media for health-related information and communication in health care: preferences of the Dutch general population. J Med Internet Res 2013 Oct 02;15(10):e220 [FREE Full text] [doi: 10.2196/jmir.2607] [Medline: 24088272]
145. Leafman JS, Mathieson K. Perceptions of telemedicine for patient education among online support group patients with chronic or rare conditions. Cogent Medicine 2018 Oct 12;5(1):1525148. [doi: 10.1080/2331205x.2018.1525148]
146. Duplaga M. Searching for a role of nursing personnel in developing landscape of eHealth: factors determining attitudes toward key patient empowering applications. PLoS One 2016;11(4):e0153173 [FREE Full text] [doi: 10.1371/journal.pone.0153173] [Medline: 27049525]
147. von Rosen AJ, von Rosen FT, Tinnemann P, Müller-Riemenschneider F. Sexual health and the internet: cross-sectional study of online preferences among adolescents. J Med Internet Res 2017 Nov 08;19(11):e379 [FREE Full text] [doi: 10.2196/jmir.7068] [Medline: 29117927]
148. Kristan J, Suffoletto B. Using online crowdsourcing to understand young adult attitudes toward expert-authored messages aimed at reducing hazardous alcohol consumption and to collect peer-authored messages. Transl Behav Med 2015 Mar;5(1):45-52 [FREE Full text] [doi: 10.1007/s13142-014-0298-4] [Medline: 25729452]
149. Wentink MM, VAN Bodegom-Vos L, Brouns B, Arwert HJ, Vlieland TP, DE Kloet AJ, et al. What is important in e-health interventions for stroke rehabilitation? A survey study among patients, informal caregivers and health professionals. Int J Telerehabil 2018;10(1):15-28 [FREE Full text] [doi: 10.5195/ijt.2018.6247] [Medline: 30147840]
150. Hughes CM, Padilla A, Hintze A, Raymundo TM, Sera M, Weidner S, et al. Developing an mHealth app for post-stroke upper limb rehabilitation: feedback from US and Ethiopian rehabilitation clinicians. Health Informatics J 2020 Jun;26(2):1104-1117 [FREE Full text] [doi: 10.1177/1460458219868356] [Medline: 31566456]
151. Ihmhemed SM, Kamaruddin RA, Atan R. Using empathy approach to design type-2 diabetic user persona. J Advanced Sci Eng Res 2017;7(1):27-42.
152. Verhoeven F, Steehouder MF, Hendrix RM, Van Gemert-Pijnen JE. From expert-driven to user-oriented communication of infection control guidelines. Int J Human Comput Stud 2010 Jun;68(6):328-343. [doi: 10.1016/j.ijhcs.2009.07.003]
153. Lambert JD, Greaves CJ, Farrand P, Haase AM, Taylor AH. Development of a web-based intervention (eMotion) based on behavioural activation to promote physical activity in people with depression. Mental Health Physical Activity 2017 Oct;13:120-136. [doi: 10.1016/j.mhpa.2017.10.003]
154. Gude WT, Simon AC, Peute LW, Holleman F, Hoekstra JB, Peek N, et al. Formative usability evaluation of a web-based insulin self-titration system: preliminary results. Stud Health Technol Inform 2012;180:1209-1211. [Medline: 22874403]
155. King S, Boutilier JA, MacLaren Chorney J. Managing chronic pain in the classroom: development and usability testing of an eHealth educational intervention for educators. Canadian J School Psychol 2016 Oct 23;33(2):95-109. [doi: 10.1177/0829573516674308]
156. Verhoeven F, Steehouder M, Hendrix R, van Gemert-Pijnen JE. How nurses seek and evaluate clinical guidelines on the internet. J Adv Nurs 2010 Jan;66(1):114-127. [doi: 10.1111/j.1365-2648.2009.05175.x] [Medline: 20423439]
157. Scholten M, Kelders SM, van Gemert-Pijnen J. A scoped review of the potential for supportive virtual coaches as adjuncts to self-guided web-based interventions. In: Persuasive Technology: Development and Implementation of Personalized Technologies to Change Attitudes and Behaviors. Cham: Springer; 2017.
158. Kip H, Bouman YH, Kelders SM, van Gemert-Pijnen JE. eHealth in treatment of offenders in forensic mental health: a review of the current state. Front Psychiatry 2018;9:42 [FREE Full text] [doi: 10.3389/fpsyt.2018.00042] [Medline: 29515468]
159. Asbjørnsen RA, Smedsrød ML, Solberg Nes L, Wentzel J, Varsi C, Hjelmesæth J, et al. Persuasive system design principles and behavior change techniques to stimulate motivation and adherence in electronic health interventions to support weight loss maintenance: scoping review. J Med Internet Res 2019 Jun 21;21(6):e14265 [FREE Full text] [doi: 10.2196/14265] [Medline: 31228174]
160. Scholten MR, Kelders SM, Van Gemert-Pijnen JE. Self-guided web-based interventions: scoping review on user needs and the potential of embodied conversational agents to address them. J Med Internet Res 2017 Nov 16;19(11):e383 [FREE Full text] [doi: 10.2196/jmir.7351] [Medline: 29146567]
161. Verhoeven F, Tanja-Dijkstra K, Nijland N, Eysenbach G, van Gemert-Pijnen L. Asynchronous and synchronous teleconsultation for diabetes care: a systematic literature review. J Diabetes Sci Technol 2010 May 01;4(3):666-684 [FREE Full text] [doi: 10.1177/193229681000400323] [Medline: 20513335]
162. Lentferink AJ, Oldenhuis HK, de Groot M, Polstra L, Velthuijsen H, van Gemert-Pijnen JE. Key components in eHealth interventions combining self-tracking and persuasive eCoaching to promote a healthier lifestyle: a scoping review. J Med Internet Res 2017 Aug 01;19(8):e277 [FREE Full text] [doi: 10.2196/jmir.7288] [Medline: 28765103]
163. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012 Nov 14;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
164. Verhoeven F, van Gemert-Pijnen JE, Dijkstra K, Nijland N, Seydel E, Steehouder M. The contribution of teleconsultation and videoconferencing to diabetes care: a systematic literature review. J Med Internet Res 2007 Dec 14;9(5):e37 [FREE Full text] [doi: 10.2196/jmir.9.5.e37] [Medline: 18093904]

165. Walters LE, Mars M, Scott RE. A review and critique of teledermatology in the South African public health sector. Stud Health Technol Inform 2016;231:143-151. [Medline: 27782026]
166. Organ D, Dick S, Hurley C, Heavin C, Linehan C, Dockray S, et al. A systematic review of user-centred design practices in illicit substance use interventions for higher education students. In: Proceedings of the European Conference on Information Systems 2018: Beyond Digitization-Facets of Socio-Technical Change. 2018 Presented at: European Conference on Information Systems 2018: Beyond Digitization-Facets of Socio-Technical Change; Jun 23 - 28, 2018; Portsmouth, UK.
167. Radomski A, Wozney L, McGrath P, Huguet A, Hartling L, Dyson M, et al. Design and delivery features that may improve the use of internet-based cognitive behavioral therapy for children and adolescents with anxiety: a realist literature synthesis with a persuasive systems design perspective. J Med Internet Res 2019 Feb 05;21(2):e11128 [FREE Full text] [doi: 10.2196/11128] [Medline: 30720436]
168. Norris A, Martinez S, Labaka L, Madanian S. Disaster e-health: a new paradigm for collaborative healthcare in disasters. In: Proceedings of the ISCRAM 2015 - Information Systems for Crisis Response and Management. 2015 Presented at: ISCRAM 2015 - Information Systems for Crisis Response and Management; May 24-27, 2015; Krystiansand, Norway. [doi: 10.13140/RG.2.1.3428.6248]
169. Meiland F, Innes A, Mountain G, Robinson L, van der Roest H, García-Casal JA, et al. Technologies to support community-dwelling persons with dementia: a position paper on issues regarding development, usability, effectiveness and cost-effectiveness, deployment, and ethics. JMIR Rehabil Assist Technol 2017 Jan 16;4(1):e1 [FREE Full text] [doi: 10.2196/rehab.6376] [Medline: 28582262]
170. Spanakis EG, Santana S, Tsiknakis M, Marias K, Sakkalis V, Teixeira A, et al. Technology-based innovations to foster personalized healthy lifestyles and well-being: a targeted review. J Med Internet Res 2016 Jun 24;18(6):e128 [FREE Full text] [doi: 10.2196/jmir.4863] [Medline: 27342137]
171. Wahle F, Bollhalder L, Kowatsch T, Fleisch E. Toward the design of evidence-based mental health information systems for people with depression: a systematic literature review and meta-analysis. J Med Internet Res 2017 May 31;19(5):e191 [FREE Full text] [doi: 10.2196/jmir.7381] [Medline: 28566267]
172. Kramer LL, Ter Stal S, Mulder BC, de Vet E, van Velsen L. Developing embodied conversational agents for coaching people in a healthy lifestyle: scoping review. J Med Internet Res 2020 Feb 06;22(2):e14058 [FREE Full text] [doi: 10.2196/14058] [Medline: 32022693]
173. Stara V, Santini S, Kropf J, D'Amen B. Digital health coaching programs among older employees in transition to retirement: systematic literature review. J Med Internet Res 2020 Sep 24;22(9):e17809. [doi: 10.2196/17809]
174. Wentzel M, Karreman J, van GJ, van VanGemert-Pijne JE, Ossebaard H, Hamalainen P. Towards an internet-based infectious disease management platform to increase patient safety. In: Proceedings of the Third International Conference on eHealth, Telemedicine, and Social Medicine. 2011 Presented at: Third International Conference on eHealth, Telemedicine, and Social Medicine; Feb 23-28, 2011; Guadeloupe, France.
175. Van Gemert-Pijnen JE, Kelders SM, Bohlmeijer ET. Understanding the usage of content in a mental health intervention for depression: an analysis of log data. J Med Internet Res 2014 Jan 31;16(1):e27 [FREE Full text] [doi: 10.2196/jmir.2991] [Medline: 24486914]
176. Sieverink F, Kelders S, Poel M, van Gemert-Pijnen L. Opening the black box of electronic health: collecting, analyzing, and interpreting log data. JMIR Res Protoc 2017 Aug 07;6(8):e156 [FREE Full text] [doi: 10.2196/resprot.6452] [Medline: 28784592]
177. Sewitch MJ, Fallone CA, Ghali P, Lee GE. What patients want in a smartphone app that supports colonoscopy preparation: qualitative study to inform a user-centered smartphone app. JMIR Mhealth Uhealth 2019 Jul 02;7(7):e12242 [FREE Full text] [doi: 10.2196/12242] [Medline: 31125310]
178. Keizer J, Braakman-Jansen LM, Kampmeier S, Köck R, Al Naiemi N, Te Riet-Warning R, et al. Correction to: cross-border comparison of antimicrobial resistance (AMR) and AMR prevention measures: the healthcare workers' perspective. Antimicrob Resist Infect Control 2019;8:133 [FREE Full text] [doi: 10.1186/s13756-019-0589-0] [Medline: 31413824]
179. Van Velsen L, Wentzel J, Van Gemert-Pijnen JE. Designing eHealth that matters via a multidisciplinary requirements development approach. JMIR Res Protoc 2013 Jun 24;2(1):e21 [FREE Full text] [doi: 10.2196/resprot.2547] [Medline: 23796508]
180. LeRouge C, Ma J, Sneha S, Tolle K. User profiles and personas in the design and development of consumer health technologies. Int J Med Inform 2013 Nov;82(11):e251-e268. [doi: 10.1016/j.ijmedinf.2011.03.006] [Medline: 21481635]
181. Tschimmel K. Design thinking as an effective toolkit for innovation. In: Proceedings of the XXIII ISPIM Conference: Action for Innovation: Innovating from Experience. 2012 Presented at: XXIII ISPIM Conference: Action for Innovation: Innovating from Experience; Jun 17-20, 2012; Barcelona, Spain.
182. Kip H. The added value of eHealth: improving the development, implementation and evaluation of technology in treatment of offenders. University of Twente. URL: https://research.utwente.nl/en/publications/the-added-value-of-ehealth-improving-the-development-implementati [accessed 2022-01-20]
183. Patrick K, Hekler EB, Estrin D, Mohr DC, Riper H, Crane D, et al. The Pace of Technologic Change. American Journal of Preventive Medicine 2016 Nov;51(5):816-824. [doi: 10.1016/j.amepre.2016.05.001]

184. Morrison LG. Theory-based strategies for enhancing the impact and usage of digital health behaviour change interventions: a review. Digit Health 2015 Jul 17;1:2055207615595335 [FREE Full text] [doi: 10.1177/2055207615595335] [Medline: 29942544]
185. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010 Feb 17;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
186. Moller AC, Merchant G, Conroy DE, West R, Hekler E, Kugler KC, et al. Applying and advancing behavior change theories and techniques in the context of a digital health revolution: proposals for more effectively realizing untapped potential. J Behav Med 2017 Feb;40(1):85-98 [FREE Full text] [doi: 10.1007/s10865-016-9818-7] [Medline: 28058516]
187. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
188. Riley WT, Rivera DE, Atienza AA, Nilsen W, Allison SM, Mermelstein R. Health behavior models in the age of mobile interventions: are our theories up to the task? Transl Behav Med 2011 Mar;1(1):53-71 [FREE Full text] [doi: 10.1007/s13142-011-0021-7] [Medline: 21796270]
189. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: a scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
190. Lorenz P, Schindler L, Steudte-Schmiedgen S, Weidner K, Kirschbaum C, Schellong J. Ecological momentary assessment in posttraumatic stress disorder and coping. An eHealth study protocol. Eur J Psychotraumatol 2019;10(1):1654064 [FREE Full text] [doi: 10.1080/20008198.2019.1654064] [Medline: 31528269]
191. Kip H, Beerlage-de JN, Wentzel J. The contextual inquiry. In: eHealth Research, Theory and Development: A Multi-Disciplinary Approach. London: Routledge; 2018.
192. Schulz K, Altman D, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. Trials 2010 Mar 24;11:32 [FREE Full text] [doi: 10.1186/1745-6215-11-32] [Medline: 20334632]
193. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
194. COREQ (Consolidated Criteria for Reporting Qualitative Studies). In: Guidelines for Reporting Health Research: A User's Manual. Hoboken, New Jersey, United States: Wiley; 2014.
195. CMD Methods Pack: find a combination of research methods that suit your needs. CMD Methods Pack. URL: https://cmdmethods.nl/ [accessed 2022-01-20]
196. HCD Design Kit California. Design Kit. URL: https://www.designkit.org/ [accessed 2022-01-19]

Abbreviations

BCT: behavior change technique
CeHRes: Center for eHealth Research
CONSORT: Consolidated Standards of Reporting Trials
COPD: chronic obstructive pulmonary disease
EMA: ecological momentary assessment
EHEC: enterohemorrhagic Escherichia coli
ICT: information and communication technology
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
VR: virtual reality

Edited by R Kukafka; submitted 07.07.21; peer-reviewed by VDB Sanne W, K Baltrusaitis, L van Velsen; comments to author 23.09.21; revised version received 29.10.21; accepted 15.11.21; published 27.01.22.

Please cite as:
Kip H, Keizer J, da Silva MC, Beerlage-de Jong N, Köhle N, Kelders SM
Methods for Human-Centered eHealth Development: Narrative Scoping Review
J Med Internet Res 2022;24(1):e31858
URL: https://www.jmir.org/2022/1/e31858
doi:10.2196/31858
PMID:35084359

Review

Implementation Frameworks for Artificial Intelligence Translation Into Health Care Practice: Scoping Review

Fábio Gama[1,2], PhD; Daniel Tyskbo[3], PhD; Jens Nygren[3], Prof Dr; James Barlow[4], Prof Dr; Julie Reed[3], Prof Dr; Petra Svedberg[3], Prof Dr

[1]School of Business, Innovation and Sustainability, Halmstad University, Halmstad, Sweden

[2]School of Administration and Economic Science, Santa Catarina State University, Florianópolis, Brazil

[3]School of Health and Welfare, Halmstad University, Halmstad, Sweden

[4]Centre for Health Economics and Policy Innovation, Imperial College Business School, London, United Kingdom

Corresponding Author:
Fábio Gama, PhD
School of Business, Innovation and Sustainability
Halmstad University
Kristian IV:s väg 3
Halmstad, 30118
Sweden
Phone: 46 0702628937
Email: fabio.gama@hh.se

Abstract

Background: Significant efforts have been made to develop artificial intelligence (AI) solutions for health care improvement. Despite the enthusiasm, health care professionals still struggle to implement AI in their daily practice.

Objective: This paper aims to identify the implementation frameworks used to understand the application of AI in health care practice.

Methods: A scoping review was conducted using the Cochrane, Evidence Based Medicine Reviews, Embase, MEDLINE, and PsycINFO databases to identify publications that reported frameworks, models, and theories concerning AI implementation in health care. This review focused on studies published in English and investigating AI implementation in health care since 2000. A total of 2541 unique publications were retrieved from the databases and screened on titles and abstracts by 2 independent reviewers. Selected articles were thematically analyzed against the Nilsen taxonomy of implementation frameworks, and the Greenhalgh framework for the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) of health care technologies.

Results: In total, 7 articles met all eligibility criteria for inclusion in the review, and 2 articles included formal frameworks that directly addressed AI implementation, whereas the other articles provided limited descriptions of elements influencing implementation. Collectively, the 7 articles identified elements that aligned with all the NASSS domains, but no single article comprehensively considered the factors known to influence technology implementation. New domains were identified, including dependency on data input and existing processes, shared decision-making, the role of human oversight, and ethics of population impact and inequality, suggesting that existing frameworks do not fully consider the unique needs of AI implementation.

Conclusions: This literature review demonstrates that understanding how to implement AI in health care practice is still in its early stages of development. Our findings suggest that further research is needed to provide the knowledge necessary to develop implementation frameworks to guide the future implementation of AI in clinical practice and highlight the opportunity to draw on existing knowledge from the field of implementation science.

(J Med Internet Res 2022;24(1):e32215) doi:10.2196/32215

KEYWORDS

implementation framework; artificial intelligence; scoping review

Introduction

Background

Artificial intelligence (AI) can potentially transform health care data into meaningful and actionable insights [1]; however, AI has not yet become widespread in health care practice. This gap in translation from research to practice is largely owing to the challenges in the implementation of AI [2,3]. This paper aims to assess the current state of academic knowledge relating to the implementation of AI and to appraise the extent to which this knowledge draws from and contrasts with knowledge about general health care technology implementation.

The potential benefits to patients from new health technologies are often missed owing to slow and variable uptake in practice [4]. The emergent field of implementation science has generated many insights into the barriers to and facilitators of the effective uptake and deployment of new health technologies [5], recognizing the unique challenges of intervening in complex health care systems compared with other industrial settings. Typical outputs from implementation science are generalizable implementation theories, models, and frameworks that are developed to describe the factors influencing implementation, predict the conditions required for successful implementation, and provide guidance for conducting and evaluating implementation efforts in health care [5-7]. These frameworks provide a general understanding of the challenges of introducing novel technologies in health care settings.

Importantly, for the development of AI technologies, implementation science has evidenced that passive approaches to the dissemination and diffusion of health care technologies are rarely effective [8]. Instead, purposive implementation efforts are required to mainstream innovation within an organization or health care system [9]. However, to date, most of the research literature on AI in health care deals with the development, application, and evaluation of advanced analytic techniques and models [10-12], primarily within computer science, engineering, and medical informatics. The literature on the implementation of AI to improve existing clinical workflows is more fragmented and mostly based on nonempirical data from proof-of-concept studies [1,13] across multiple subject areas, such as data governance [14], ethics [15], accountability [3], interpretability [16], and regulation [17]. This means that there are uncertainties around factors that influence the implementation of AI in real-world health care setups [10] and that health care professionals lack guidance on how to implement AI in their daily practices [18].

If the value of AI technologies is to be realized in practice, it is important to develop evidence-based approaches to AI implementation. Although it is likely that generalizable implementation theories, models, and frameworks will be able to provide valuable guidance for the implementation of AI technologies, it is likely that the nature of AI features will add new layers of complexity and pose additional challenges to effective implementation [2]. First, AI differs in its potential to augment or constrain the work of health care professionals compared with other technologies. This difference shifts attention away from predicting successful implementations of *passive* technologies (eg, telehealth or pedant alarms) to understand how health care professionals and AI interact to create value for patients and other users. Second, AI challenges our dichotomic belief that divides the realms of human aptitudes and machine capabilities. Recent AI developments allow the perception of emotions, conversations, and, ultimately, creativity. Such capabilities allow AI to enter into domains that were previously exclusive to humans. Third, AI implementation is highly complex, requiring activities that cover a wide range of stakeholders from technology developers, system regulators, organizations and individuals, professionals, patients, and caregivers. This puts AI at the more complex end of what has been studied by implementation sciences, which tend to be well-defined and bounded interventions. Combined, these differences suggest the need to develop an AI-specific evidence base for implementation, and for generalizable knowledge about AI implementation to be shared in AI-specific implementation theories, models, and frameworks.

Objectives

This study aims to explore the current state of academic knowledge of AI implementation by assessing any implementation theories, frameworks, or models that are specific to AI translation into health care practice. The study objectives are to assess the following:

1. What, if any, AI-specific implementation frameworks in health care exist?
2. How do these AI-specific implementation frameworks draw on and compare to more generalized implementation frameworks for health technologies?
3. What do any AI-specific implementation frameworks reveal about the challenges of AI implementation?

Methods

Study Design

An interpretative scoping review was considered the most appropriate method to answer the research questions, as it provides a systematic synthesis of knowledge within a defined area, and with the aim of exploring and mapping key concepts, available evidence, and shortcomings in existing research [19,20]. The following steps were taken: (1) deciding on definitions, (2) systematically searching in databases, (3) selecting and screening studies, and (4) extracting and analyzing the data.

Deciding on Definitions

The operational definitions for the term implementation, framework, and AI are listed in Table 1. These broad terms were used for searches to cover any literature, specifically using the terms in combination. Using broad terms for the searches instead of more specialized exemplifying terms, the review aims to cover all literature on implementation frameworks for AI translation into health care practice that is framed within these broad terms and thus could be interpreted as generally applicable to this context. Our operational definition for implementation was inspired by Greenhalgh et al [9], which refers to an “active and planned effort to mainstream an innovation within an organization.” According to this definition, implementation

efforts are purposeful and are described in sufficient detail such that independent observers can recognize the presence and strength of active and planned actions. Although other definitions of implementation exist, we chose this definition, as it is broad and enables us to identify multiple elements from prior AI-related studies.

Table 1. Operational definitions for key concepts.

Term	Operational definition	Examples in health care
Implementation	An intentional effort designed to change or adapt or uptake interventions into routines [9].	• Adoption of heart failure prediction software • Change the clinical decision support system
Artificial intelligence	A general purpose technology based on a core set of capabilities and computational algorithms designed to mimic human cognitive functions to analyze complex data [2].	• Machine learning for mortality prediction • Unstructured image data analysis for radiology
Framework	A simplification structure, overview, system or plan of multiple descriptive categories or elements (ie, constructs, concepts, and variable) that streamline the interpretation of a phenomenon [21].	• NASSS[a] framework (for health and care technologies) [5] • SHIFT[b] evidence [7]

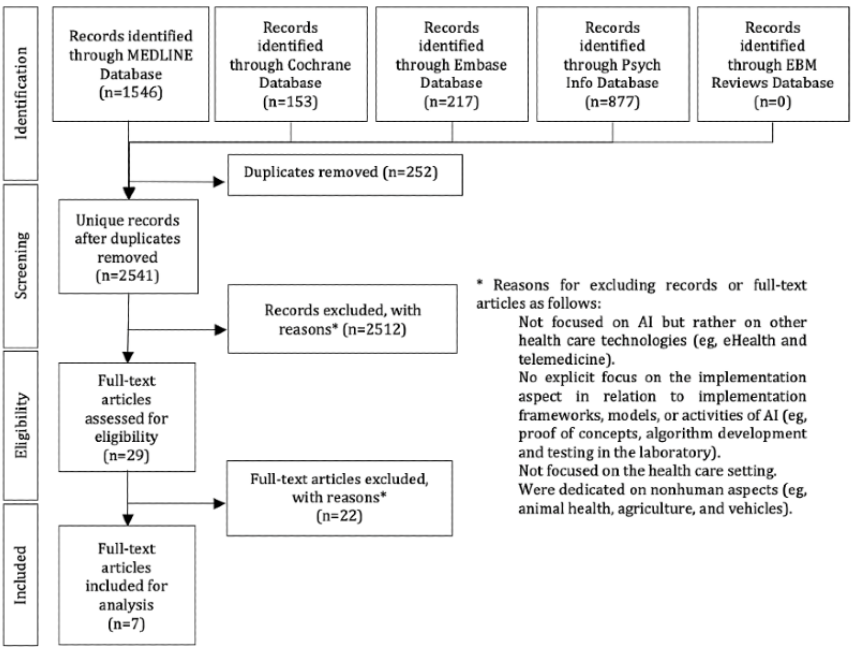
[a]NASSS: nonadoption, abandonment, scale-up, spread, and sustainability.

[b]SHIFT: successful health care improvement from translating evidence.

Systematically Searching in Databases

A systematic search of MEDLINE, Embase, EBM *Reviews*, PsycINFO, and Cochrane databases was performed following the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram (Figure 1). The search was limited to the literature published from January 1, 2000, to March 1, 2020, and all original publications that described AI-specific implementation frameworks based on health care studies. Search terms included a combination of terms relating to implementation (*implementat**), frameworks (*model* and *theory*), AI (*artificial intelligence*), and machine learning (*ML*; Multimedia Appendix 1). Several different terms could potentially be used as synonymous with the term AI and included as search terms. However, as the purpose of this study was to explore the literature on AI-specific implementation frameworks, our starting point was to include only those studies that presented such frameworks specifically in relation to the overarching concepts of AI and ML. A similar strategy was used by Wolff et al [22] to define the search terms. The MeSH (Medical Subject Headings) vocabulary was used to accurately define the search terms, and the MeSH terms supervised and unsupervised ML were both captured by the term *machine learning*. This strategy was motivated by preliminary literature reviews on AI, consultation with AI experts, and librarians.

Figure 1. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flow diagram showing the review process. AI: artificial intelligence; EBM: Evidence Based Medicine.



Selecting and Screening Studies

Studies were eligible for inclusion if they were written in English and referred to the implementation of AI (eg, ML) in health care settings. All study designs or publication types were eligible for inclusion to identify the presence of implementation frameworks to guide the use of AI in health care. Conference abstracts, editorials, and technical reports were excluded. Other reasons for exclusion included studies that did not focus on AI, had no explicit focus on implementation in relation to implementation frameworks, models, or theories of AI, that were not focused on the health care setting, or were dedicated to nonhuman aspects (eg, animal health). Titles and abstracts were screened for inclusion by 2 independent reviewers (FG and DT) using the Rayyan web platform. Disagreements were resolved by consensus, and when necessary, a third reviewer was involved (JB). The agreement score during screening was substantial (κ score>0.8).

Extracting and Analyzing the Data

A 4-step process was used for data extraction and analysis according to the analytical framework of Arksey et al [19]. First, 4 reviewers (FG, DT, JN, and PS) independently piloted a structured data extraction tool on the same final included articles. The reviewers discussed and compared their analyses, and any disagreements were resolved via discussion and consensus. Following the guidelines by Arksey et al [19], information related to country, study design, area of practice, target population, study focus, study aims, and any literature cited as informing the framework development was extracted to understand the main areas of interest.

Second, as this field of research is relatively new, we considered it important to perform a quality assessment of the included articles, even if this is not typical for scoping reviews [19,23]. Thus, we have deliberately addressed concerns that the lack of quality assessment in scoping reviews makes the interpretation and translation of the findings more challenging [24]. The quality of the selected studies was assessed using the Critical Appraisal Skills Programme (CASP) [25]. Two authors (FG and PS) independently rated the articles and resolved disagreements through consensus (Multimedia Appendix 2 [26-32]). The CASP appraisal checklist includes multiple study designs and was therefore chosen as the most appropriate tool to evaluate the selected articles. Articles were not excluded owing to poor quality, as suggested by other studies [33]. Instead, we considered the relative contribution from high and low-quality studies in the analysis phase when discussing the presence of implementation frameworks and elements for AI implementation.

Third, to thematically analyze the literature, the frameworks used in the included studies were categorized in accordance with the Nilsen taxonomy of 5 categories of implementation frameworks [21]. These categories include process models, determinant frameworks, classic theories, implementation theories, and evaluation frameworks. This conceptual foundation is important, as it helps to understand the challenges associated with implementation. Process models aim to guide or describe the process of translating research into practices (eg, stages and phases). Determinant frameworks intend to explain or understand the influence of variables on implementation outcomes (eg, barriers and enablers). Determinant frameworks account for 5 types of determinants that focus on the characteristics of individual elements, including objects (eg, AI), users or adopters, end user (eg, patients), context, and strategy. Both classic theories and implementation theories are distinguished from research-to-practice models, although they explain how change occurs without ambitions to bring about the change. The differences between these 2 theories are that classic theories have been developed from a field external to implementation science (ie, psychology and sociology) and implementation theories have been developed by researchers in the field of implementation science. What all theories have in common is that they have attempted to have some predictive capacity and explain causal mechanisms. Evaluation frameworks specify aspects of implementation to evaluate and determine implementation success (eg, checklists and criteria). To further strengthen the analysis even further, the included studies were categorized as either including a formal (explicitly integrated body of knowledge) or informal framework (implicit assumptions, beliefs, and views), all following the definitions provided by Nilsen [21].

Fourth, to analyze the literature in relation to the extent to which these findings draw on existing implementation frameworks, a deductive thematic analysis was carried out. This entailed coding the included studies deductively according to the nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework for health technologies [5]. This framework suggests a range of subdomains that are relevant to understanding how a health technology might be implemented ranging from a specific health condition to a wider regulatory and sociocultural system. It represents a state-of-the-art structure and has been used in other AI studies [2]. Five reviewers (FG, DT, JN, JR, and PS) interpreted how implementation elements captured in the descriptive themes were related across the NASSS domains. The NASSS framework domains are the condition, technology, value proposition, adopters, organization, wider system, and embedding and adaptation over time. This involved extensive reflection on article findings and how the findings were related to the domains. The process began by deductively assessing the findings in each article and then evaluating their association with NASSS domains. An analytic summary matrix was developed by tabulating the domains for each of the included studies within a table, including identification of new subdomains that did not fit within the existing NASSS framework. Finally, the implementation elements were summarized across the NASSS subdomains and newly identified subdomains to assess the inclusion of elements across the included articles.

Results

The Characteristics of Published Literature

The initial search returned 2541 unique articles. We screened all abstracts and eliminated 98.86% (2512/2541) of the papers based on the exclusion criteria. The term implementation framework was widely dispersed across different areas of health care, which resulted in a high number of nonrelated papers.

After abstract screening, 29 articles were subjected to full-text review, of which 76% (n=22) of them did not meet the inclusion criteria. Finally, 24% (7/29) of articles were included.

Out of 7 articles, only 2 (29%) of the articles included formal frameworks that directly addressed AI implementation in health care. Owing to the limited number of articles fulfilling the inclusion criteria, articles that partially met the criteria were also included: 14% (1/7) of the articles that included a formal framework addressing issues of ethics and AI (a topic of high importance to real-world AI implementation), and 57% (4/7) of the articles that included descriptions of elements influencing implementation (focusing on physician opinion, patient opinion, and factors influencing the use of AI in emergency departments and surgical settings).

The articles presented heterogeneous study designs. In total, 57% (4/7) of these studies were literature reviews, 29% (2/7) used a qualitative approach, and 71% (1/14) used quantitative survey data. The area of practice and the focus of the selected articles were dispersed. The papers focused on the perceptions of patients and clinicians of AI (n=2), decision-making and decision support systems (n=2), ethical and trustworthy aspects (n=1), benefits and challenges in implementing AI (n=1), and implementation elements to guide AI adoption (n=1). Further characteristics of the included studies are shown in Table 2.

Table 2. Characteristics of included papers (n=7).

Study	Country	Study design	Area of practice	Target population	Study focus	Study aims
Beil et al [26]	Germany	Literature review	Intensive care	N/A[a]	Ethical and trustworthy aspects in intensive care	Discuss ethical considerations about AI[b] for prognostication in intensive care.
Diprose et al [27]	New Zealand	Quantitative study	Primary care	Physicians (n=170)	Perceptions of clinicians to understanding, explain and trust on AI results	Investigate the association between physician understanding of AI outputs, their ability to explain these to patients, and their willingness to trust the AI outputs.
Fernandes et al [28]	Portugal	Literature review	Emergency department	N/A	Intelligent CDSS[c] for triage	Assess how intelligent CDSS for triage have been contributing to the improvement of quality of care in the ED[d] as well as to identify the challenges they have been facing regarding implementation.
Loftus et al [29]	United States	Literature review	Operation room	Surgeons	Decision-making in surgeries	Propose that AI models would obviate these weaknesses and be integrated with bedside assessment to augment surgical decision-making.
Nelson et al [30]	United States	Qualitative study	Dermatology clinics	Patients from general dermatology clinics (n=48)	Perception of patients on AI related to skin cancer screening	Explore how patients conceptualize AI and perceive the use of AI for skin cancer screening.
Ngiam et al [31]	Singapore	Literature review	Health care	N/A	Benefits and challenges of AI in oncology	Discuss some of the benefits and challenges of big data and machine learning in health care.
Truong et al [32]	Canada	Qualitative study	Health care	Subject-matter experts in health care (n=8)	Implementation elements to guide AI adoption	Creating an implementation framework to help health care organizations understand the key considerations and guide implementation efforts for AI.

[a]N/A: not applicable.

[b]AI: artificial intelligence.

[c]CDSS: clinical decision support system.

[d]ED: emergency department.

Quality Assessment

Overall, the quality assessment indicated that 43% (3/7) of articles [26,31,32] met below 30% of the CASP criteria owing to limited descriptions of the methodological procedures. In total, 43% (3/7) of the articles [27-29] met between 50% and 60% of the criteria, whereas 14% (1/7) of articles [30] met 70% of the criteria. The quality scores of each paper in relation to the CASP checklist are provided in the (Multimedia Appendix 2).

Framework Categories to Implement AI

The analysis identified the use of three primary framework categories: determinant framework, process model, and evaluation framework. Classic theories and implementation theories have not yet been identified. Across the 7 articles, the data analysis indicated that 3 (43%) articles included explicit frameworks: determinant frameworks [32], process models [31], and evaluation frameworks [26]. The remaining articles (4/7, 57%) did not include comprehensive frameworks; instead, elements relevant to implementation could be identified. For example, Beil et al [26] discussed the applicability of ethical constructs in an AI implementation, and Diprose et al [27] explored physician opinion of implementation elements, whereas Nelson et al [30] suggested implementation elements from the patient perspective. These articles did not identify or describe any conceptual or organizational framework to illustrate the relationship between implementation elements. The types of frameworks and elements of each framework (eg, determinants, steps, or aspects) are provided in Table 3. Except for 29% (2/7) of the articles [31,32], most articles lacked a clear description of the elements identified (Table 3).

The data analysis identified the presence of implementation elements in all 7 of the NASSS domains (Table 4). The implementation elements were unequally distributed across the domains and subdomains. A full description of the implementation elements identified in each paper and their relation to NASSS subdomains are provided in the data analysis matrix (Multimedia Appendix 3 [26-32]). Across all papers, technology was the most frequently included domain identified in all articles (7/7, 100%), while embedding and adaptation over time was the less frequently included domain (1/7, 14%). Collectively, the papers explicitly covered 91% (20/22) of the NASSS subdomains; however, there was an uneven distribution between subdomains and between individual articles. Across the 7 articles, the subdomains most frequently identified were *material and features of technology* (7/7, 100%), *types of data generated* (7/7, 100%), and *staff (role and identity*, 6/7, 86%). Less frequently identified subdomains include *scope for adaptation over time* (1/7, 14%) and *extent of change needed to routines* (1/7, 14%).

Among the individual articles, Nelson et al [30] and Beil et al [26] had the highest coverage with 68% (15/22) and 55% (12/22) of the subdomains, respectively. Diprose et al [27] and Fernandes et al [28] covered the fewest subdomains, with only 23% (5/22) of the subdomains. The presence of the elements was not mutually exclusive, as some elements were classified into multiple domains. For example, Beil et al [26] classified the implementation element called explicability in the domains technology, adopters, and wider systems. Similarly, the implementation element human-machine interaction from Ngiam et al [31] was classified in the technology, adopters, organization, and wider system.

In total, 7 new subdomains were identified that did not explicitly fit in the NASSS framework (Table 4). Three elements were classified as belonging to the technology domain: *types of data input*, *dependence/adaptation to the local context*, and *evaluation of effectiveness*. The *types of data inputted* highlight the essential data input into AI algorithms and their dependence on the availability and quality of existing (automated) electronic health record data [29]. *Dependence/adaptation to local contexts* relates to how AI technologies are dependent on other care practices that vary at a local level, such as other technologies in use, data recording methods, and local process of care (eg, patient referral systems) that influence the data input to AI or the ability of local practitioners to trust and act on AI outputs [28]. *Evaluation of effectiveness* makes explicit the need for clinical trials and other validation mechanisms to ascertain the effectiveness, reliability, and trustworthiness of AI algorithms [32]. Under the value proposition domain, we identified the *demand-side value (to population)*, which refers to a medical ethical principle that requires fairness and societal well-being in AI implementation at the population level and considers issues such as data bias and implications for health inequalities within the value proposition of the technology [26]. For domain adopters, we identified *shared decision-making* among patients, professionals, caregivers, and the role of AI as a fourth voice within decision-making processes [30]. Finally, for the domain wider system, we identified *ethics (population equity/discrimination)* and the *role of human oversight*. *Ethics* indicates medical moral principles for beneficence and nonmaleficence at a population level, reflecting the need for formal regulation and consideration of the health equality impact of new technologies at the population level [26,29]. *Human oversight* considers the extent to which it is possible or desirable for technologies to operate with or without human oversight. Together, these underline that AI should not undermine the autonomy of health care professionals nor provoke adverse effects [26,29].

Table 3. Descriptions of the frameworks and framework elements in the included articles (n=7).

Study	Explicit framework?	Types of framework[a] and purpose	Framework elements (stages, determinants, or aspects)	Clarity of element description[b]	Referenced guidance or literature for framework development
Beil et al [26]	Yes	Evaluation framework; ethical AI[c]	Beneficence, nonmaleficence, justice, autonomy, explicability, medical perspective, technical requirements, patient- or family-centered, and system-centered	Partial	European Commission guideline
Diprose et al [27]	No	N/A[d]; elements describe physician opinion of AI	Physician understanding and intended physician behavior, explainability, preferred to explainability methods	Partial	Absent
Fernandes et al [28]	No	N/A; elements describe limitations to develop and implementing AI in ED[e] triage	Availability of data, the subjectivity of the system, methodologies and modeling techniques, validation, and geography (data from the same geographic area)	Partial	Absent
Loftus et al [29]	No	N/A; elements describe challenges and potential of AI in surgical decision-making	Challenges in surgical decision-making (complexity, values and emotions, time constraints and uncertainty, heuristics and bias), traditional predictive analytics and clinical decision support (decision aids and prognostic scoring systems), AI predictive analytics and augmented decision-making (machine learning, deep learning, and reinforcement learning), implementation (automated electronic health record data, mobile device outputs, and human intuition), challenges to adoption (safety and monitoring, data standardization and technology infrastructure, interpretability, and ethical challenges)	Partial	Absent
Nelson et al [30]	No	N/A; elements describe patient opinion of AI	AI concept, AI benefits, AI risks, AI strengths, AI weaknesses,[f] AI implementation (symbiosis, credibility, diagnostic tool, setting, and integration into electronic health records. Challenges include malpractice, misunderstanding of AI, and regulations), response to conflict between human and AI clinical decision-making, responsibility for AI accuracy, responsibility for AI data privacy, AI recommendation	Limited	Absent
Ngiam et al [31]	Yes	Process model; AI development and implementation	Clinical problem definition or redefinition, data extraction selection, and refining, data analysis and validation, human-machine interaction, paper trial, prospective clinical trial, medical device registration, and clinical deployment	Explicit	Absent
Truong et al [32]	Yes	Determinant framework; AI implementation	Data quality and quantity, trust, ethics, readiness for change, expertise, buy-in (value creation), regulatory strategy, scalability and evaluation	Explicit	Absent

[a]Type of framework according to the Nilsen taxonomy [21].

[b]Explicit: explicit definition; partial: some discussion, but no explicit definition; limited: only listed construct names, but no definition or discussion is provided.

[c]AI: artificial intelligence.

[d]N/A: not applicable.

[e]ED: emergency department.

[f]Only categories associated with artificial intelligence implementation are shown in full.

Table 4. A comparison of elements identified in literature with the nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability (NASSS) framework domains. (n=7).

Condition	Technology	Value proposition	Adopters	Organization	Wider system	Embedding and adaptation over time
Original NASSS subdomains						
• Nature of condition[a] (n=0) • Comorbidities, sociocultural influences (n=0)	• Material and features of technology (n=7) • Types of data generated (n=7) • Knowledge needed to use (n=5) • Technology supply model (n=2)	• Supply-side value (to developer; n=2) • Demand-side value (to patient; n=2)	• Staff (role and identity; n=6) • Patient (simple vs complex input; n=3) • Carer (available, nature of input; n=2)	• Capacity to innovate (n=1) • Readiness for change (n=2) • Nature of adoption or funding decision (n=1) • Extent of change to new routines (n=1) • Work needed to implement change (n=2)	• Political or policy (n=2) • Regulatory or legal (n=5) • Professional (n=2) • Sociocultural (n=2)	• Scope for adaptation over time (n=1) • Organizational resilience (n=1)
New NASSS subdomains						
• Not identified	• Types of data inputted (n=3) • Dependence on other local processes and practices (n=2) • Evaluation of effectiveness (n=3)	• Demand-side value (to population; n=1)	• Shared decision-making (n=3)	• Not identified	• Ethics (population equity or discrimination; n=2) • Role of human oversight[b] (n=3)	• Not identified

[a]These elements were not explicitly mentioned in the framework or list of elements, but they were considered in the manuscript (nature of condition, 6 articles; comorbidities and sociocultural influences, 2 articles).

[b]Can be considered across multiple domains.

Discussion

Principal Findings

This literature review demonstrates that understanding how to implement AI in health care practice is still in its early stages of development. Although our study search terms identified a large number of articles, only 7 articles were included in the final analysis. Only 29% (2/7) of these articles included formal frameworks that directly addressed AI implementation in health care, and the other 71% (5/7) of the articles provided descriptions of elements influencing such implementation.

The importance of developing knowledge of how to implement AI in health care was highlighted in many of the rejected articles, but despite acknowledging this, the articles provided little or no substance to support their claims, or guidance on how to move forward. A challenge to building knowledge in this field was underscored during the screening process, where many articles mentioned AI but were excluded because they focused on health care technologies unrelated to AI; for example, eHealth and telemedicine. The inappropriate labeling of technologies as AI likely reflects the hype surrounding the AI concept and the tendency to adopt fashionable terms to increase attention, readership, and likeliness of publication [34,35]. This type of misuse of AI terminology creates a murky landscape and ambiguity for researchers attempting to synthesize learning in this emerging field.

Given the recognition of the importance and challenges of AI implementation [2], it was surprising to find that none of the identified articles referred to existing implementation literature in informing data analysis or framework development. Although AI is likely to have specific requirements compared with other health technologies, there is a wealth of literature on implementation challenges and facilitators (eg, Xiang et al [36]) that could inform the AI field and accelerate learning. The only paper informing framework development [26] was the Ethics Guidelines for Trustworthy AI by the European Commission. This suggests that there are attempts within the AI community to reach agreement on key issues relating to the real-world use of AI and that these efforts have not yet been connected with insights from the implementation science community.

To understand the overlap of concerns between AI-specific implementation challenges and more general health

technologies, we mapped the elements listed in the 7 papers against the NASSS framework, which was specifically developed to guide the uptake of health technologies [5]. Most elements identified in the papers corresponded to factors within the NASSS framework, suggesting that there is a significant degree of overlap between the concerns of AI implementation and general health technologies, although it should be noted that the general lack of clarification of the element descriptions at times meant they were open to interpretation. The NASSS domain most identified in the 7 AI papers was technology, including the subdomain *material features of the technology* and *types of data generated*, which were included in all 7 papers. The importance of *adopters (staff roles)* was most frequently mentioned (6/7, 86%), followed by *wider system regulatory or legal issues* (5/7, 71%). All other NASSS domains and subdomains were identified in 1 or 2 other papers (except for the condition domain, which did not formally appear in any framework or list of elements, but was nonetheless part of the contextual discussion in most papers). These findings suggest that all NASSS elements are relevant to the implementation of AI in health care, but that awareness and recognition of all these domains are currently low within the AI community. This highlights the value of sharing findings within the AI community, and between AI and implementation science communities, to build a more comprehensive understanding.

A small number of elements identified in the 7 papers did not align with the existing NASSS subdomains. As such, we propose 7 new subdomains that can supplement the NASSS framework. Of the newly identified subdomains, some highlight issues that are likely relevant to all forms of health technologies and may have specific implications within AI. For example, the newly identified subdomain of *shared decision-making* recognizes the need for processes and behaviors that support communication, discussion, and decision-making among staff, patients, and careers (technology adopters), which is likely to be relevant for many data-driven technologies [37-39]. However, in the case of AI, the AI provides a fourth *voice* in the decision-making process that will have particular implications for how such communication is handled in an emotionally sensitive manner, how much weight is given to different opinions and preferences [38,40], and how it could support clinical decision-making without compromising the primary responsibilities and duties of the health care professional for patient care [41]. Similarly, the need for the *evaluation of effectiveness* is important for all health technologies [42,43], but for AI, this may be of particular importance in demonstrating the trustworthiness of data outputs if it is to replace or complement clinical judgment.

The subdomain *role of human oversight* has emerged as a unique implementation feature across multiple domains. The term *human oversight* underlines that AI should not undermine the autonomy of health care professionals nor provoke adverse effects [26,29]. Patients are receptive to the use of AI, yet health care professionals need to have oversight of the AI outcomes and decide when and how to use the information generated from the AI. For example, in particular cases, health care professionals might be able to override the decision made by AI. Unlike *shared decision-making*, which suggests collaborative work between humans and machines for individual patient care, *human oversight* underscores the mandate of health care professionals over AI recommendations as a critical element at the regulatory and system levels. This subdomain has been echoed in the gray literature under analogous terms such as human-in-the-loop, human-on-the-loop, or human-in-command [44].

Other newly identified subdomains appear to be more highly relevant or potentially problematic for AI than for other forms of health technology. For example, understanding the *demand-side* value (population benefit) and *ethics (population equity or discrimination)* appear to be of particular importance to AI, given their dependence or impact on large (population size) data. The possibility of recommending decisions across an entire population entails risks to reinforce systemic biases (eg, White or male), which might unintentionally discriminate minorities and patients with more complex or unusual health conditions. This, in turn, is linked to the increased importance of regulatory and legal systems that oversee the introduction of AI and carefully consider the implications and responsibilities for individuals, professional groups, and governments to ensure the safety, effectiveness, ethics, and equity of new data-driven technologies [40].

Implications for Practice and Research

Our findings highlight the need to develop an AI-specific implementation framework, drawing on empirical research related to AI implementation efforts, and drawing on existing knowledge and experience within the implementation science community. There is a great, currently unrealized, opportunity to draw on insights from the implementation of science literature to enhance and accelerate the implementation of AI. There is no need to repeat the mistakes or reproduce learning that has already been achieved, and there is an opportunity to use the theoretical and practical insights from others to provide an evidence-based foundation that can accelerate the implementation of AI in health care.

Our findings suggest that the implementation of AI is viewed through a narrow lens, focusing on the design of the technology and its interaction with the immediate user. Lessons from implementation science suggest the need to extend attention to understand how the technology will influence and interact with the context in which it is implemented, including understanding existing processes and practices of care within each local setting, and how systems work at micro, meso, and macro levels to support or hinder technology uptake. Such insights can only be gained from active engagement of relevant stakeholders, frontline staff, patients, and careers, their engagement is essential to understand how new technologies that are based on AI will be received, and how trust can be built to ensure that design is centered on the needs and practical constraints and requirements of the health care system (ie, to produce useful, trusted, relevant, and actionable knowledge). Lessons from implementation science suggest that obtaining this knowledge, using it to inform technology design, and addressing wider implementation issues is time intensive and reliant on good quality relationships between diverse and often conflicting groups of stakeholders. However, time and time again the implementation literature demonstrates the necessity of this work for successful

implementation. Harnessing such insights could provide guidance to health care professionals responsible for implementing AI in practice, for decision makers and policy makers to ensure effective implementation plans are in place, and for the AI designers and promotors who need to be aware of the implications of real-world deployment to ensure that AI products are suitable for implementation.

Any AI implementation framework also needs to recognize the heightened and perhaps unique needs and challenges of introducing AI in health care, including meaningful decision support, ethical dilemmas (privacy and consent), transparency, effectiveness, interpretability, and establishing trust in *black box* technologies [2,45,46]. Implementation is further complicated by the symbiotic relationship between AI and the system, the dynamic and interdependent relationship between data input from the health care system to inform the AI, as well as the influence of AI data output on the same system, and the significant technical, analytic, and clinical expertise required to understand and resolve problematic issues. In addition, AI has a greater potential and risk of operating at the population level and the ethical and regulatory requirements to ensure that any such technologies provide equitable population care and safe, effective, and compassionate care at an individual level.

Methodological Considerations

Although this study was conducted in a structured and systematic manner, only a small number of papers met the inclusion criteria, and these papers were of rather low quality in terms of methodological clarity and rigor, and in the clarity of descriptions and definitions of elements influencing AI implementation. In addition, the included studies that were based on empirical data were conducted only in 3 high-income countries, limiting the generalization of the findings to other contexts. For example, the use of AI in health care in low-income countries is still nascent, and therefore some subdomains of the NASSS framework might be irregularly advanced in this context (eg, legal, regulatory, and social cultural). These characteristics emphasize the importance of reflecting on the findings of parsimony. Together, these limit the reliability and generalizability of the cumulative findings from our analysis and highlight a gap in the literature that requires further empirical and theoretical research.

We chose to use the NASSS framework for deductive analysis of the included papers as it represents one of the most advanced frameworks dedicated to understanding implementation of health care technologies and was informed by extensive empirical research and literature review [2]. Although the use of the NASSS framework has provided a helpful way of making sense of the findings from the AI implementation literature, the use of this framework is illustrative and other frameworks could have been applied to support review and interpretation of results. For example, the NASSS framework adopts a predominantly innovator-centric view of understanding how new technologies can be introduced into a health care system, whereas alternative frameworks provide a more service-centric view in understanding the needs and operating reality of frontline health care services and how new technologies fit with and disturb existing working practices [47]. Further research could be conducted to consider the merits and limitations of different implementation frameworks to provide insights into AI implementation.

Scoping reviews often search for the identification and conceptualization of complex, emergent, or ill-defined concepts. Unlike traditional systematic reviews guided by well-defined constructs, it may be unfeasible to screen and synthesize all relevant literature on an emergent topic [48]. As our purpose was to merely understand what implementation frameworks have been used in the application of AI for health care practices, our efforts to identify all eligible studies were limited in some respects. Moreover, although the results of this review are based on 7 studies, the empirical data were gathered solely from 5 health care–orientated databases. Although a focused sampling technique reduces contextual variations and thus helps provide robust findings, health care professionals working outside health care contexts should consider context-specific variations as they interpret the findings.

Conclusions

This literature review demonstrates that the research literature on AI implementation in health care lacks theoretical development and is poorly connected to existing implementation frameworks or models developed within implementation science. This means that potential specific challenges around AI implementation are largely unrevealed, and that further empirically based research is needed to provide the knowledge necessary to develop implementation frameworks to guide future implementation of AI in clinical practice.

Acknowledgments

The funders for this study are the Swedish Government Innovation Agency Vinnova (grant 2019-04526) and the Knowledge Foundation (grant 20200208 01H). The funders were not involved in any aspects of study design, collection, analysis, interpretation of data, or in the writing or publication process.

Authors' Contributions

All authors (FG, DT, JN, JR, JB, and PS) made significant contributions to the original paper. All the authors contributed to the study design. Applications for funding and coproduction agreements were put in place by PS and JN. FG, DT, and JB identified and selected relevant studies, FG and DT charted the data, and FG, DT, JN, JR, and PS analyzed, summarized, and reported the results. The manuscript was drafted by FG, DT, JN, JR, and PS, and all authors critically revised the paper in terms of important intellectual content. All authors have read and approved the final submitted version.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Health care database search syntax.
[DOCX File , 31 KB - jmir_v24i1e32215_app1.docx]

Multimedia Appendix 2
Quality appraisal of the selected papers.
[DOCX File , 36 KB - jmir_v24i1e32215_app2.docx]

Multimedia Appendix 3
Data analysis matrix.
[DOCX File , 97 KB - jmir_v24i1e32215_app3.docx]

References

1. Mehta N, Pandit A, Shukla S. Transforming healthcare with big data analytics and artificial intelligence: a systematic mapping study. J Biomed Inform 2019;100:103311. [doi: 10.1016/j.jbi.2019.103311] [Medline: 31629922]
2. Shaw J, Rudzicz F, Jamieson T, Goldfarb A. Artificial intelligence and the implementation challenge. J Med Internet Res 2019;21(7):e13659. [doi: 10.2196/13659] [Medline: 31293245]
3. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. Nat Med 2019;25(1):30-36. [doi: 10.1038/s41591-018-0307-0] [Medline: 30617336]
4. Whitelaw S, Pellegrini D, Van Spall HG. Barriers to and facilitators of the uptake of digital health technology in cardiology: a systematic review. Eur Heart J 2020;41(Supplement_2):ehaa946.3500. [doi: 10.1093/ehjci/ehaa946.3500]
5. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
6. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
7. Reed JE, Howe C, Doyle C, Bell D. Successful healthcare improvements from translating evidence in complex systems (SHIFT-evidence):simple rules to guide practice and research. Int J Qual Health Care 2019;31(3):238-244. [doi: 10.1093/intqhc/mzy160] [Medline: 30085160]
8. Lyon AR, Lewis CC. Designing health information technologies for uptake: development and implementation of measurement feedback systems in mental health service delivery. Introduction to the special section. Adm Policy Ment Health 2016;43(3):344-349. [doi: 10.1007/s10488-015-0704-3] [Medline: 26658691]
9. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
10. Alhashmi SF, Alshurideh M, Al Kurdi B, Salloum SA. A systematic review of the factors affecting the artificial intelligence implementation in the health care sector. In: Hassanien AE, Azar A, Gaber T, Oliva D, Tolba F, editors. Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020). New York City: Springer; 2020:37-49.
11. Horgan D, Romao M, Morré SA, Kalra D. Artificial intelligence: power for civilisation – and for better healthcare. Public Health Genomics 2019;22(5-6):145-161. [doi: 10.1159/000504785] [Medline: 31838476]
12. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
13. Ben-Israel D, Jacobs WB, Casha S, Lang S, Ryu WH, de Lotbiniere-Bassett M, et al. The impact of machine learning on patient care: a systematic review. Artif Intell Med 2020;103:101785. [doi: 10.1016/j.artmed.2019.101785] [Medline: 32143792]
14. Lam K, Iqbal FM, Purkayastha S, Kinross JM. Investigating the ethical and data governance issues of artificial intelligence in surgery: protocol for a Delphi study. JMIR Res Protoc 2021;10(2):e26552. [doi: 10.2196/26552] [Medline: 33616543]
15. Murphy K, Di Ruggiero E, Upshur R, Willison DJ, Malhotra N, Cai JC, et al. Artificial intelligence for good health: a scoping review of the ethics literature. BMC Med Ethics 2021;22(1):14. [doi: 10.1186/s12910-021-00577-8] [Medline: 33588803]
16. Wongvibulsin S, Wu KC, Zeger SL. Improving clinical translation of machine learning approaches through clinician-tailored visual displays of black box algorithms: development and validation. JMIR Med Inform 2020;8(6):e15791. [doi: 10.2196/15791] [Medline: 32515746]

17. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020;22(6):e15154. [doi: 10.2196/15154] [Medline: 32558657]
18. Esmaeilzadeh P. Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives. BMC Med Inform Decis Mak 2020;20(1):170 [FREE Full text] [doi: 10.1186/s12911-020-01191-1] [Medline: 32698869]
19. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. Int J Soc Res Methodol 2005;8(1):19-32. [doi: 10.1080/1364557032000119616]
20. Colquhoun HL, Levac D, O'Brien KK, Straus S, Tricco AC, Perrier L, et al. Scoping reviews: time for clarity in definition, methods, and reporting. J Clin Epidemiol 2014;67(12):1291-1294. [doi: 10.1016/j.jclinepi.2014.03.013] [Medline: 25034198]
21. Nilsen P. Making sense of implementation theories, models and frameworks. Implement Sci 2015;10:53 [FREE Full text] [doi: 10.1186/s13012-015-0242-0] [Medline: 25895742]
22. Wolff J, Pauling J, Keck A, Baumbach J. The economic impact of artificial intelligence in health care: systematic review. J Med Internet Res 2020;22(2):e16866. [doi: 10.2196/16866] [Medline: 32130134]
23. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 2018;169(7):467-473. [doi: 10.7326/M18-0850] [Medline: 30178033]
24. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
25. Nadelson S, Nadelson LS. Evidence-based practice article reviews using CASP tools: a method for teaching EBP. Worldviews Evid Based Nurs 2014;11(5):344-346. [doi: 10.1111/wvn.12059] [Medline: 25156640]
26. Beil M, Proft I, van Heerden D, Sviri S, van Heerden PV. Ethical considerations about artificial intelligence for prognostication in intensive care. Intensive Care Med Exp 2019;7(1):70. [doi: 10.1186/s40635-019-0286-6] [Medline: 31823128]
27. Diprose WK, Buist N, Hua N, Thurier Q, Shand G, Robinson R. Physician understanding, explainability, and trust in a hypothetical machine learning risk calculator. J Am Med Inform Assoc 2020;27(4):592-600. [doi: 10.1093/jamia/ocz229] [Medline: 32106285]
28. Fernandes M, Vieira SM, Leite F, Palos C, Finkelstein S, Sousa JM. Clinical decision support systems for triage in the emergency department using intelligent systems: a review. Artif Intell Med 2020;102:101762. [doi: 10.1016/j.artmed.2019.101762] [Medline: 31980099]
29. Loftus TJ, Tighe PJ, Filiberto AC, Efron PA, Brakenridge SC, Mohr AM, et al. Artificial intelligence and surgical decision-making. JAMA Surg 2020;155(2):148-158. [doi: 10.1001/jamasurg.2019.4917] [Medline: 31825465]
30. Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, et al. Patient perspectives on the use of artificial intelligence for skin cancer screening: a qualitative study. JAMA Dermatol 2020;156(5):501-512. [doi: 10.1001/jamadermatol.2019.5014] [Medline: 32159733]
31. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019;20(5):e262-e273. [doi: 10.1016/s1470-2045(19)30149-4] [Medline: 31044724]
32. Truong T, Gilbank P, Johnson-Cover K, Ieraci A. A framework for applied AI in healthcare. Stud Health Technol Inform 2019;264:1993-1994. [doi: 10.3233/SHTI190751] [Medline: 31438445]
33. Thomas J, Harden A. Methods for the thematic synthesis of qualitative research in systematic reviews. BMC Med Res Methodol 2008;8:45 [FREE Full text] [doi: 10.1186/1471-2288-8-45] [Medline: 18616818]
34. Dwivedi YK, Hughes L, Ismagilova E, Aarts G, Coombs C, Crick T, et al. Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int J Inf Manage 2021;57:101994. [doi: 10.1016/j.ijinfomgt.2019.08.002]
35. Fröhlich H, Balling R, Beerenwinkel N, Kohlbacher O, Kumar S, Lengauer T, et al. From hype to reality: data science enabling personalized medicine. BMC Med 2018;16(1):150. [doi: 10.1186/s12916-018-1122-7] [Medline: 30145981]
36. Xiang Y, Zhao L, Liu Z, Wu X, Chen J, Long E, et al. Implementation of artificial intelligence in medicine: status analysis and development suggestions. Artif Intell Med 2020;102:101780. [doi: 10.1016/j.artmed.2019.101780] [Medline: 31980086]
37. Begley K, Begley C, Smith V. Shared decision‐making and maternity care in the deep learning age: acknowledging and overcoming inherited defeaters. J Eval Clin Pract 2021;27(3):497-503. [doi: 10.1111/jep.13515] [Medline: 33188540]
38. Bjerring JC, Busch J. Artificial intelligence and patient-centered decision-making. Philos Technol 2020;34(2):349-371. [doi: 10.1007/s13347-019-00391-6]
39. Elwyn G, Nelson E, Hager A, Price A. Coproduction: when users define quality. BMJ Qual Saf 2020;29(9):711-716. [doi: 10.1136/bmjqs-2019-009830] [Medline: 31488570]
40. Gooding P, Kariotis T. Ethics and law in research on algorithmic and data-driven technology in mental health care: scoping review. JMIR Ment Health 2021;8(6):e24668. [doi: 10.2196/24668] [Medline: 34110297]
41. Lysaght T, Lim HY, Xafis V, Ngiam KY. AI-assisted decision-making in healthcare: the application of an ethics framework for big data in health and research. Asian Bioeth Rev 2019;11(3):299-314. [doi: 10.1007/s41649-019-00096-0] [Medline: 33717318]
42. Mueller D, Tivey D, Croce D. Health-technology assessment: its role in strengthening health systems in developing countries. South Afr J Public Health 2017;2(1):6-11. [doi: 10.7196/SHS.2017.v2.i1.50]

43. Vis C, Bührmann L, Riper H, Ossebaard HC. Health technology assessment frameworks for eHealth: a systematic review. Int J Technol Assess Health Care 2020;36(3):204-216. [doi: 10.1017/s026646232000015x] [Medline: 32297588]
44. High-Level Expert Group on Artificial Intelligence. Ethics guidelines for trustworthy AI. European Commission, Directorate-General for Communications Networks, Content and Technology. Brussels: Publications Office of the European Union; 2019. URL: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419 [accessed 2021-07-10]
45. Wirtz BW, Weyerer JC, Geyer C. Artificial intelligence and the public sector—applications and challenges. Int J Public Adm 2019;42(7):596-615. [doi: 10.1080/01900692.2018.1498103]
46. Benbya H, Pachidi S, Jarvenpaa S. Special issue editorial: artificial intelligence in organizations: implications for information systems research. J Assoc Inf Syst 2021;22(2):10. [doi: 10.17705/1jais.00662]
47. Reed JE, Green S, Howe C. Translating evidence in complex systems: a comparative review of implementation and improvement frameworks. Int J Qual Health Care 2019;31(3):173-182. [doi: 10.1093/intqhc/mzy158] [Medline: 30060185]
48. Pham MT, Rajić A, Greig JD, Sargeant JM, Papadopoulos A, McEwen SA. A scoping review of scoping reviews: advancing the approach and enhancing the consistency. Res Synth Methods 2014;5(4):371-385. [doi: 10.1002/jrsm.1123] [Medline: 26052958]

Abbreviations

AI: artificial intelligence
CASP: Critical Appraisal Skills Programme
MeSH: Medical Subject Headings
NASSS: nonadoption, abandonment, scale-up, spread, and sustainability
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses

Edited by G Eysenbach; submitted 19.07.21; peer-reviewed by S Nilsson, M Fernandes, H Hah; comments to author 11.08.21; revised version received 02.09.21; accepted 27.12.21; published 27.01.22.

Please cite as:
Gama F, Tyskbo D, Nygren J, Barlow J, Reed J, Svedberg P
Implementation Frameworks for Artificial Intelligence Translation Into Health Care Practice: Scoping Review
J Med Internet Res 2022;24(1):e32215
URL: https://www.jmir.org/2022/1/e32215
doi:10.2196/32215
PMID:35084349

Review

Virtual Reality Simulation for Disaster Preparedness Training in Hospitals: Integrated Review

Younhyun Jung[1], DPhil

School of Computing, Gachon University, Seongnam-si, Republic of Korea

Corresponding Author:
Younhyun Jung, DPhil
School of Computing
Gachon University
1332, Seongnam-daero
Sujeong-gu
Seongnam-si, 13120
Republic of Korea
Phone: 82 317504771
Email: younhyun.jung@gachon.ac.kr

Abstract

Background: A critical component of disaster preparedness in hospitals is experiential education and training of health care professionals. A live drill is a well-established, effective training approach, but cost restraints and logistic constraints make clinical implementation challenging, and training opportunities with live drills may be severely limited. Virtual reality simulation (VRS) technology may offer a viable training alternative with its inherent features of reproducibility, just-in-time training, and repeatability.

Objective: This integrated review examines the scientific evidence pertaining to the effectiveness of VRS and its practical usefulness in training health care professionals for in-hospital disaster preparedness.

Methods: A well-known 4-stage methodology was used for the integrated review process. It consisted of problem identification, a literature search and inclusion criteria determination, 2-stage validation and analysis of searched studies, and presentation of findings. A search of diverse publication repositories was performed. They included Web of Science (WOS), PubMed (PMD), and Embase (EMB).

Results: The integrated review process resulted in 12 studies being included. Principle findings identified 3 major capabilities of VRS: (1) to realistically simulate the clinical environment and medical practices related to different disaster scenarios, (2) to develop learning effects on increased confidence and enhanced knowledge acquisition, and (3) to enable cost-effective implementation of training programs.

Conclusions: The findings from the integrated review suggested that VRS could be a competitive, cost-effective adjunct to existing training approaches. Although the findings demonstrated the applicability of VRS to different training scenarios, these do not entirely cover all disaster scenarios that could happen in hospitals. This integrated review expects that the recent advances of VR technologies can be 1 of the catalysts to enable the wider adoption of VRS training on challenging clinical scenarios that require sophisticated modeling and environment depiction.

(J Med Internet Res 2022;24(1):e30600) doi:10.2196/30600

KEYWORDS

virtual reality; in-hospital disaster preparedness training; mass casualty incidents; hospitals

Introduction

Emergency preparedness in hospitals for disasters is unexpectedly required. Disasters could occur anywhere and at any time [1,2]. Disasters include not only accidents in hospitals (eg, an outbreak of fire) but also natural and human-made catastrophic events occurring outside of hospitals. In-hospital disaster preparedness must be performed fully on the spot, and health care professionals are the front line of disaster preparedness by controlling disaster situation and caring for victim patients. Human history has demonstrated the serious effects of poor disaster preparedness [3]. In the case of the aftereffect of Hurricane Katrina, the emergency evacuation at the Memorial Medical Center in New Orleans was unexpectedly

conducted and poor evacuation management resulted in tremendous loss of life [4].

A fundamental component of effective in-hospital disaster preparedness is appropriate, repetitive experiential education and training of health care professionals with live disaster situations [5]. Disasters, however, are rarely encountered clinically, making it ideal for simulation. Live drills enable trainees to practice in a realistic simulation, where they can experience mock-up scenarios and learn sequences of actions and tasks to deal with the situations. Live drills are available when disaster simulations are set through the process of implementing everything realistically, including participants, infrastructures, and medical devices and tools. One of the live drill examples is the timely and effective evacuation practice of neonates during an outbreak of fire [6]. Here, fast exit routing by newborn intensive care unit workers while caring for mannequins of neonates had to be developed. Live drills have a great advantage in that they can take effect in a real disaster situation. Cost restraints and logistic constraints, however, make implementing simulations and scenarios difficult, and training opportunities with live drills may be severely limited [7,8].

Having lectures and practice with learning materials is another approach. Its advantages are that it is simple and does not require time and cost on a large scale. The simulations and instructions, however, may be fragmentary (eg, via verbal description or simple figures), and the learning outcomes may not be so effective in a real disaster situation [9].

Emerging evidence suggests that virtual reality simulation (VRS) is a viable alternative for workforce training in a variety of industry disciplines, with positive effects of increasing confidence and gaining necessary knowledge [10-13]. VR is a cutting-edge technology that can generate virtual environments of different training scenarios and combine with multisource data, stereoscopic vision, and intuitive interaction interfaces [14]. VR enables trainees to take part in virtual scenarios using avatars, in which they can simulate the sequences of actions and tasks they may make and reflect on the consequences of their choices [15,16]. Its realistic and interactive characteristics can greatly enhance trainees' perception level and interest in learning [17]. The current generation of VR technologies is now equipped with head-mounted displays, where the wearers can obtain fully immersive experience of VRS [18]. There are various advantages of the use of VRS in education and training as follows: (1) repeatability, where training practices can be repetitive until goals are achieved [19]; (2) dynamic expansion and update of training scenarios due to technical advances to reconstruct a variety of training scenarios in a realistic manner [20,21]; and (3) just-in-time training, where training can be conducted anywhere and at any time if VR devices are available [22].

The purpose of this paper is to conduct an integrative review of peer-reviewed literature that applies VRS for in-hospital disaster preparedness training. The research problems this integrated review identified are follows: For use in in-hospital disaster preparedness, (1) *what kinds of VRS training programs have been presented?*; (2) *are there any studies that demonstrate the usefulness of the VRS training in a quantitatively or qualitatively manner, and does this usefulness analysis include technological feasibility, training effects, and cost saving of VRS?*; and (3) *are there any studies that identify the pros and cons of VRS training when compared to conventional alternatives (eg, live drills or material-based lectures)?*

This new integrated review builds upon and complements a prior integrated review publication by Miller et al [23]. In 2013, they analyzed peer-reviewed literature published during the period of 2005-2012, with findings that the efficacy of VRS for its use in disaster preparedness training was identified but such that the findings were difficult to be generalized due to the small volume of the literature. This new integrated review analyzed state-of-the-art literature published in the past 15 years, where there have been massive technological advances in VR engineering [17]. This motivated us to investigate the current progress in disaster preparedness training programs using VRS in terms of its wide adoption, practical usages, and efficiency. Compared to the prior publication [23], this new integrated review narrowed down the application domains and was only limited to in-hospital events and scenarios (eg, emergency departments) where health care professionals should equip patient triage and treatment skills for mass casualty incidents (MCIs) [24], which are critical and should be timely and carefully managed.

Methods

Study Design

This study followed the well-known methodology suggested by Whittemore et al [25] to conduct an integrative review. This review methodology consists of (1) problem identification, (2) a literature search and inclusion criteria definition, (3) validation of searched studies, and (4) presentation of findings. Two investigators—an assistant professor (author YJ) and a senior student of computer science—were involved in this study.

Problem Identification

The investigators agreed on the purpose and boundaries of this study at the initial stage of the research meeting. It seems that health care professionals do not have enough opportunities for in-hospital disaster preparedness training [7,8]. The investigators also agreed on recent advances of VRS for its use in education and training in a variety of application industry domains [10-13]. The research problems mentioned in the Introduction section were then derived based on their agreement.

Literature Search and Inclusion Criteria Definition

The literature search was performed by the investigators independently. They used 3 keywords: (1) "virtual reality," which represents a key technology to be used for implementation of in-hospital disaster preparedness training programs; (2) "disaster" or "fire," which represents catastrophic events to be simulated for training programs; and (2) "health care" and "hospital," which represent major responders when the training programs are executed. The combination of these keywords can cover the volume of relevant literature for this integrated review. Note that the chosen keywords—"disaster" and "fire"—sufficiently covered different possible disaster types, where "disaster" represents natural and human-made

catastrophic events (eg, earthquake and MCIs) and "fire" represents accidents the majority of the literature considers as catastrophic events occurring in hospitals.

A computerized search approach was used with 3 websites: Web of Science (WOS), PubMed (PMD), and Embase (EMB). WOS is an established publication repository for VR technologies (technology domains), and the other 2 are related to the medicine application domain. Such inclusion may decrease the possibility of missing relevant literature. The advanced search function of each publication repository was used as a search option as follows:

- The WOS search with its Web of Science Core Collection database and without any restrictions on document types
- The PMD search with all fields and without any other options or filters
- The EMB search with its default options

Inclusion criteria were as follows:

- Peer-reviewed empiric studies published from 2006 to 2020 and written in English
- All VRS training programs related to disaster preparedness
- Participants of training programs being all types of health care professionals
- Place of training programs anywhere in hospitals

Validation of Searched Studies

The final studies to be included were determined through a 2-step validation procedure: (1) Abstracts of searched studies were initially reviewed to confirm whether they met the inclusion criteria, and (2) full papers of the abstract-filtered studies were printed and read in their entirety for inclusion. Many of the studies have contributed to the development of VRS training programs, and there was a lack of evaluating and analyzing their usefulness and efficacy in real practice scenarios. These were excluded since the purpose of this integrated review was to suggest the scientific evidence pertaining to the practical values of VRS training programs. The investigators independently performed the validation procedure. Discrepancies were resolved through active discussion among the investigators.

Presentation of Findings

A matrix of the presentation for validated final studies was used to reach an integrated result. The matrix consisted of authors and dates of publication, study design, disaster type, participant details, contents, and outcomes. The investigators believed that this categorization would be useful to capture and promote the understanding of features of VRS disaster preparedness training programs and their usefulness and practical values. For quasi-experimental studies, it is important to identify differences and similarities, describe variables, and clarify intervention factors to make logical connections. All of these were carefully considered in this integrated review.

Results

Study Selection

There were 99 results (studies) from the literature search of the 3 publication repositories, as shown in Table 1. After study screening using the inclusion criteria and redundancy removal, 12 (12%) studies in total were included in this integrated review: 4 (24%) studies were included from 17 studies in WOS, 6 (26%) from 23 studies in PMD, and 2 (3%) from 59 studies in EMB.

Table 1. Numerical results of the literature search.

Search results	WOS[a], n (%)	PMD[b], n (%)	EMB[c], n (%)	Total
Studies after initial literature search	17 (17.2)	23 (23.2)	59 (59.6)	99
Studies after screening using inclusion criteria	7 (32)	6 (27)	9 (41)	22
Final studies after redundancy removal	4 (33)	6 (50)	2 (17)	12

[a]WOS: Web of Science.

[b]PMD: PubMed.

[c]EMB: Embase.

Study Review

The summarization for each of the 12 studies [6,24,26-35] is shown in Table 2 in ascending order of publication year. Here, there were 4 categories: authors and publication years, experiment, content, and outcomes. In the experiment category, data including the status of the randomized controlled trial (RCT) and disaster type and study design were described. In addition, the number of participants was provided, with a distribution of their role—either the experiment group (EG) with VRS training or the control group (CG) without VRS training—if a quasi-experimental study was conducted. Prior to experiments, both groups were taught through a conventional training approach (eg, lecture with learning materials) and an additional VRS training session was delivered only to the EG.

In Table 3, statistical characteristics of the 12 studies [6,24,26-35] are shown according to predefined categories.

Table 2. Summary of the 12 studies.

Author (year)	Experiment	Content	Outcome
Roy et al [26]	• Case study • Bioterrorism MCIs[a] • Physicians (number not specified) • RCT[b]: N/A[c] • EG[d]: N/A • CG[e]: N/A	In VRS[f] sessions, participants were asked to conduct interviews with simulated patients and exams with virtualized resources (eg, medical images and heart sounds). Live drill simulations equivalent to the VRS sessions were also given for comparison. Qualitative and quantitative feedback was obtained to assess participants' performance, along with detailed conversational analysis. Software platform: SIMmersion	Analysis results suggested that VRS training can build skills, increase learning retention, improve trainee confidence, and change behaviors. VRS training requires greater initial investment but has lower marginal costs when compared to the live drill counterpart, and considerable return on investment.
Heinrichs et al [27]	• Case study • Bioterrorism MCIs • 7 physicians and 6 nurses • RCT: N/A • EG: N/A • CG: N/A	In VRS sessions, triage teams of participants assessed victims and sent them to appropriate treatment areas. The teams divided themselves into physician-nurse teams at each bedside and assessed and managed each virtual patient allocated. During the assessment, additional graphical information was given, and they performed intravenous infusions and administered blood and drugs. After completing the VRS sessions, an instructor facilitated a debriefing discussion. The participants were asked to conduct the same VRS sessions once more. After the second debriefing, they were asked to complete surveys and contribute to an open discussion. Most (9/13, 69%) of the participants were not gamers (69% never played VR systems), and most (8/13, 62%) had no prior training in responding to an MCI. Software platform: Online Interactive Virtual Environment (OLIVE)	Here, 8 (62%) of 13 participants reported that the VRS training changed their feelings and attitudes about working as members of an emergency department team. The ratings of the participants on the exit survey (5-point Likert-type scale: 1=low and 5=high) showed that they felt immersed (3.47) and thought that the VRS training increased their confidence in their ability to respond to MCIs (2.0 before training; 3.08 after training). Most also thought that the VRS training would be useful for learning teamwork skills and behaviors (3.77) as well as for learning the clinical skills necessary to treat MCI victims (3.15). Their comments also indicated they perceived the patient physiology models and virtual environment as realistic, although they would like the interface improved to allow them to perform a more rapid patient assessment.
Heinrichs et al [24]	• Sequential study • Bioterrorism MCIs • 10 physicians (4 years of postgraduate experience) and 12 nurses (9.5 years of practice experience) • RCT: N/A • EG: N/A • CG: N/A	Participants conducted VRS of 2 MCI sessions. For each session, they formed 2 teams—those assigned to a triage area and those assigned to an immediate treatment area—and began to act out their signed roles for assessing and treating victim patients. After the VRS sessions, the participants joined an instructor-led debriefing of their VRS performance and then filled out an exit questionnaire and contributed to a focus group discussion. Quantitative results were collected from a quiz that was administered at the beginning of the evaluation and an exit questionnaire that was completed at the end. The majority had never played VR games: the mean score on the frequency of play was 1.4 between “never” and “occasionally.” Approximately, two-thirds of the participants had previous triage training at some point prior to study enrollment. Software platform: OLIVE	Prior to the VRS training, only 4 (18%) of 22 participants were confident about managing MCIs. After the VRS training, 19 (86%) felt either “confident” or “very confident,” with 13 (59%) attributing this change to practicing in the VRS training. In addition, 21 (95%) reported that the session scenarios were useful for improving health care team skills training, and 18 (82%) believed that the sessions also were instructive in learning about clinical skill management of MCIs.
Pucher et al [28]	• Sequential study • Bomb terrorism MCIs • 21 clinicians in 3 groups: 8 novices, 7 intermediates, and 6 experts • RCT: N/A • EG: N/A • CG: N/A	In VRS, participants in each of the 3 groups were asked to form a team and required to perform clinical action to ensure appropriate place, transfer, and treatment for virtual patients. Participants were allowed to access additional information (eg, each patient's notes or vital signs). Technical skill performance of individual participants was collected on a 5-point Likert scale across a range of critical behaviors and tasks defined by a disaster planning expert panel. Nontechnical skill performance was scored based on the validated trauma nontechnical skills (T-NOTECHS). Scores were compared across groups. The participants filled a feedback and validity questionnaire, with statement responses on a 7-point Likert scale. Software platform: Unity	All 21 (100%) participants agreed that VRS would be an effective and realistic training tool for MCIs and that it was an enjoyable addition to their training and might help improve their own practice. The novice group committed more critical events than the expert group (11 novice vs 3 expert, P=.01), took longer to treat patients (560 seconds vs 399, P=.03), and resulted in poorer T-NOTECHS scores (14 vs 21.5, P=.003) and technical skill scores (2.29 vs 3.96, P=.001). Participants who previously underwent disaster response training thought that VRS has significant advantages over existing alternatives, but details of the advantages were not stated explicitly.

Author (year)	Experiment	Content	Outcome
Ferra et al [6]	• Quasi-experimental study • Bioterrorism MCIs • 106 nursing students • RCT: yes • EG: n=54 • CG: n=52	Both EG and CG completed pretests of self-efficacy and cognitive learning. The EG conducted a VRS session where participants were required to practice sequential steps of decontamination skills with virtual tools (eg, donning personal protective equipment). After the VRS session, the EG completed posttests. Both EG and CG were then directed to a mannequin, on which they demonstrated decontamination while being evaluated and timed by an experienced observer. In total, 38 (35.8%) of 106 participants had previous disaster training prior to study enrollment. Software platform: not specified	The EG reported high levels of satisfaction with VRS as a training method. A significantly shorter amount of completion time from the EG was shown when compared to the CG (P=.01). The EG showed greater improvement in the self-efficacy score over the CG, although there was no significant difference (P=.17). No difference between the EG and CG was found in the cognitive knowledge score (P=.63).
Dorozhkin et al [29]	• Sequential study • Fire • 49 physicians • RCT: N/A • EG: N/A • CG: N/A	Participants were asked to complete an operating room fire training/prevention sequence given by a VRS session. They were then asked to answer a subjective preference questionnaire (5-point Likert-type scale) focused on the usefulness and fidelity of the VRS. Software platform: VEST[g]	Five questions focusing on VRS effectiveness and its usefulness in operating room fire safety training were rated above 4. The highest score (4.84) was given to the level of satisfaction of using VRS to learn the principles rather than just using textbooks, with the lowest score (2.95) associated with the quality of sensation of feeling the tools on the target and in the task space. A total of 33 (67%) of 49 participants chose VRS training over traditional approaches, such as a textbook or an animal model.
Dubovsky et al [30]	• Sequential study • Disaster not specified • 10 nurses (25.3 years of practice experience) • RCT: N/A • EG: N/A • CG: N/A	Participants mastered navigating VRS for code triage of virtual patients in an emergency department. They then participated in a testing scenario with code triage of a series of 6 virtual patients, which represented a range of severity and complexity. Participants decided which patient was seen next based on their assessment of priority. Attitudes toward the VRS and perceived workload in the VRS and on the job were assessed with an exit questionnaire and the NASA task load index. Only 1 (10%) of 10 participants had experience with VR games. Software platform: CliniSpace	Responses to the exit questionnaire indicated that the participants' attitudes toward the VRS were largely positive. Participants generally regarded the scenarios as realistic and perceived their work on the VRS task to be equivalent to their workload in a regular workday in all aspects except for physical exertion. The time to perform code triage corresponded to the time required in the emergency department, and virtual patients were appropriately prioritized according to severity.
Ferra et al [31]	• Quasi-experimental study • Disaster not specified • 93 newborn intensive care unit health care works • RCT: yes • EG: at least 31 • CG: at least 31	A longitudinal experiment was conducted to study both the EG and CG, with repeated measures taken at 0, 4, 8, and 12 months. Learning was measured using a cognitive assessment and self-efficacy questionnaire. The EG's qualitative experience was collected using a focus group. In addition, longitudinal performance was assessed with live evacuation drills before the study and 12 months after the study. In each period, the EG was asked to conduct VRS of emergency evacuation scenarios that augmented the materials developed by an established institution. The CG was asked to review the web-based lecture materials that deliver the same content as in the VRS. Software platform: not specified	The evaluation demonstrated mixed but overall positive results for the effect of VRS on newborn intensive care unit evacuation training. The EG and CG did not statistically differ based on the scores on cognitive assessment or perceived self-efficacy. The EG performance in the live evacuation drills, however, was statistically (P<.001) and clinically (effect size of 1.71) better than that of the CG. The EG showed slightly faster transfer of neonates, but this effect did not reach statistical significance.
Sankaranarayan et al [32]	• Quasi-experimental study • Fire • 20 physicians • RCT: not specified • EG: 10 • CG: 10	Both the EG and CG took a pretest that assessed the baseline knowledge in operating room fire and its prevention. The EG was asked to practice on a VRS session of a fire scenario within a week from the pretest. In the VRS session, the EG was asked to identify the elements of the fire triangle and the proper sequence of actions that needs to be taken if a virtual patient is on fire. A week after the posttest, both groups also participated in a live drill and simulated a mock-up fire scenario, while their performance was videotaped for assessment by 2 independent raters. Software platform: VEST	Median test scores for the CG increased from 5.5 to 9.00 (P=.01) and for the EG increased from 5.0 to 8.5 (P=.01). Both groups started at the same baseline (pretest, P=.53) and reached similar levels in cognitive knowledge (posttest, P=.85). When evaluated in the live drill, 7 (70%) of the EG participants were able to perform the correct sequence of steps in extinguishing the simulated fire, whereas only 2 (20%) of the CG participants were able to do that (P=.003).
Lovreglio et al [33]			

Author (year)	Experiment	Content	Outcome
	• Sequential study • Earthquake • 87 visitors and hospital staff • RCT: yes • EG: N/A • CG: N/A	In total, 87 participants were randomly selected from 170 candidates. They were asked to practice a VRS session that was designed to generate training outcomes (ie, enhance participants' knowledge of how to behave in public and administrative areas of a hospital during and after an earthquake). The performance on the VRS session was assessed using a 6-point Liker scale questionnaire. Confirmatory factorial analysis was also run with the components forming the questionnaire. Software platform: Unity	The results from the questionnaire indicated that all its components received a positive score (eg, high rating score of 0.830 for the realism of the VRS environment). The component having the lowest score was realism of the nonplayer characters. The confirmatory factorial analysis result indicated that the realism of the virtual environment and the realism of earthquake simulation and damage play were the main contributing factors to the sense of presence available from the VRS. *U*-test results with the demographic information showed that there were no statistical differences related to the participant gender and type (staff vs visitors).
Rossler et al [34]	• Quasi-experimental study • Fire • 20 prelicensure nursing students • RCT: yes • EG: 5 • CG: 15	The EG completed a VRS training session designed for acquisition of knowledge of operating room fire safety. Both the EG and CG were then asked to complete a simulated fire live drill designed to assess the transfer of knowledge to practice settings. An investigator-developed fire safety evaluation test was administered in pretest (prior to the live drill)-posttest (after the live drill) format. Software platform: VEST	Both groups started at the same baseline in the acquisition of required knowledge (ie, the median score of 70 in the pretest). The EG showed a large increase (20 points) in gained knowledge compared with the CG (10 points), but there were no statistically significant findings for either group (between pre- and posttest).
Farra et al [35]	• Quasi-experimental study • Disaster not specified • 91 newborn intensive care unit health care workers • RCT: not specified • EG: 34 • CG: 57	A live drill was financially compared with the VRS counterpart. The costs of the live drill included exercise planning, exercise participants, exercise support, and exercise evaluation. Staff costs were based on the average hourly rate of representatives with a given title of those who were involved. The costs of the VRS included storyboard, consultants, training simulation development, travel from the development team, hardware supplies, and staff time for training participation. To have a meaningful comparison, the authors projected the cost of each alternative, assuming that all 334 staff members of the hospital were to undergo training once a year for 3 years. Software platform: Unity	The larger initial investment in the VRS can be spread across a large number of trainees and a longer time period with little additional cost, while each live drill requires additional costs that scale with the number of participants. Initially, the VRS was more expensive, with its cost of $327.78 per participant (the total cost of $106 951.14 per exercise) versus $229.79 (total cost $18 617.54) for the live drill. When development costs were extrapolated to repeated training over 3 years, however, the VRS training became less expensive, with a cost of $115.43 per participant, while the cost of live exercises remained fixed.

[a]MCI: mass casualty incident.

[b]RCT: randomized controlled trial.

[c]N/A: not applicable.

[d]EG: experiment group.

[e]CG: control group.

[f]VRS: virtual reality simulation.

[g]VEST: Virtual Electrosurgical Skill Trainer.

Table 3. Statistical characteristics of the 12 studies.

Variable	Category	Statistics, n (%)
Publication year		
	2006-2010	3 (25)
	2011-2015	2 (16)
	2016-2020	7 (59)
Study design		
	Case study design	2 (16)
	Sequential design	5 (42)
	Quasi-experimental design	5 (42)
Disaster type		
	Bioterrorism MCIs[a]	4 (34)
	Bomb terrorism MCIs	1 (8)
	Fire	3 (25)
	Earthquake	1 (8)
	Not specified	3 (25)
Participants, n		
	1-50	7 (59)
	51-100	3 (25)
	>100	1 (8)
	Not specified	1 (8)
Experiment type		
	RCT[b]	6 (50)
	No RCT	4 (34)
	Not specified	2 (16)
Software platform		
	Unity	3 (25)
	VEST[c]	3 (25)
	Online Interactive Virtual Environment (OLIVE)	2 (16)
	Others	2 (16)
	Not specified	2 (16)

[a]MCI: mass casualty incident.

[b]RCT: randomized controlled trial.

[c]VEST: Virtual Electrosurgical Skill Trainer.

Discussion

Principal Findings

All the 12 studies suggested positive outcomes of VRS from its use for in-hospital disaster preparedness training. Since the outcomes varied, the interpretation for the studies was multidimensional, focusing on 3 major themes: VRS training's (1) realism, (2) learning effect, and (3) cost saving.

Realism of VRS Training

The definition of realism was how the health care participants felt that the training environment provided by VRS was like their actual clinical environment and medical practice. There were 4 (33%) of 12 studies [27,28,30,33] where the analysis and evaluation of VRS training programs with regard to realism was explicitly conducted. These studies [27,28,30,33] demonstrated the capabilities of VR technology in realistically simulating clinical scenarios and medical practices related to disasters. The confirmatory factorial analysis by Loverglio et al [33] suggested that such the replicability is the main contributing factor in the sense of realism. Here, the health care participants could interact with sequences of actions and tasks required to deal with the in-hospital disaster scenarios [27,28,30]. These findings may indicate that VR technology

enables immersive training and learning in the clinical domain. This is consistently found in prior integrated reviews in medicine [23] as well as those in other nonmedical industry domains [12,13].

In a prior integrated review by Miller et al [23], triage and treatment training from human-made MCIs was only for application practices. This limited applicability was mitigated in the current integrated review by including new programs of building evacuation safety training from natural catastrophic events, such as earthquakes [33]. The technological advances in VRS during the past 15 years may be 1 of the contributing factors in the application to clinical scenarios that require sophisticated modeling and environment depiction.

There were complains about the VR interface from an early study by Heinrich et al [27] in 2008, but no such complaints were discovered in recent studies until 2020. This might imply that the technological maturity of VRS is at a level that includes consideration of extraneous variables, such as technological sophistication of the participants and ease of navigation.

Learning Effects of VRS Training

In total, 10 (83%) of 12 studies [6,24,26-32,34] (2 studies by Ferra et al [35] and Lovreglio et al [33]) evaluated VRS for its learning effects on in-hospital disaster preparedness training. Learning effects on increased confidence and enhanced knowledge acquisition were captured in all the 10 studies [6,24,26-32,34], and such learning effects were consistently found in previous integrated reviews in medicine [23] as well as other nonmedical industry domains [12].

In a prior integrated review by Miller et al [23], the learning effects were limited in triage and treatment knowledge and skills because all included studies focused on those practices. The larger variation of applied practices was found in our integrated review; it included not only triage practices but also training programs on fire prevention and safety, and emergency evacuation planning and execution. Through this new integrated review, it was found that the learning effects from the use of VRS can be generalized to a variation of medical training practices.

There were 4 (33%) quasi-experimental studies [6,31,32,34] where EG participants with VRS training were analyzed and directly compared with their CG counterparts without VRS training. These studies [6,31,32,34] consistently showed no significant difference between the 2 groups in posttest evaluation on self-efficacy or knowledge acquisition. The EG participants, however, showed statistical or clinical performance enhancement in gained knowledge and task completion efficiency when evaluated in final live drills (eg, performing the correct sequence of steps in extinguishing a simulated fire) [32]. This result suggested the practical usefulness of VRS and its learning effects on real disaster scenarios that health care professionals may encounter in the future.

One of the important learning effects is increased learning retention [36,37]. The prior integrated review by Miller et al [23], however, could not derive any findings on it due to a lack of studies. In this integrated review, there was 1 (8%) study [31] that took the learning retention effect into consideration. The study [31] conducted a longitudinal experiment with repeated measures taken during a year, and the results showed improved knowledge gaining as training progressed. This statement, however, is somewhat difficult to be generalized due to the insufficient volume of longitudinal experimental studies. This insufficiency may be, in part, attributed to the fact that inputs and efforts, including participants, facilities, and program management, are much greater and have to be maintained for a longer period when compared to 1-shot studies.

Cost Saving of VRS Training

Cost analysis of the training program implementation and operation is 1 of the important factors training bodies have to consider, but it was not part of the previous integrated review by Miller et al [23]. In this integrated review, 2 (16%) of 12 relevant studies [26,35] suggested consistent findings for the cost advantage in implementing and operating VRS training programs over the live drill alternative. The larger initial investment from VRS training was undeniable, but it could be scalable and compensated for its longer-term use with the large number of participants. A comparative numerical study by Farra et al [35] found that the initial additional cost from VRS training was $97.99 when compared to the live drill alternative, and this initial investment provided a benefit of $114.65 after 3 years of program operation. Farra et al [35] attributed this to the repeatability feature of VRS anywhere and anytime, enabling just-in-time training.

Other Aspects of VRS Training

This integrated review of 12 studies [6,24,26-35] showed the applicability of VRS to different training scenarios of in-hospital disaster preparedness, including triage and transport of victim patients, emergency treatment, fire extinguishing, and building evacuation. These, however, do not entirely cover all disaster scenarios that could happen in hospitals. Investigating VRS with other challenging clinical scenarios that require sophisticated environment depiction and complicated training tasks (eg, patient protection and evacuation in scenarios of water entering a hospital due to floods) is an interesting future direction for effective in-hospital disaster management. The recent maturity of VR technologies in modeling and rendering and the development tools would be able to address the complexity of different clinical scenarios.

The number of VRS studies on in-hospital disaster preparedness training programs is still limited. Compared to the previous integrated review by Miller et al [23], the number increased from 10 to 12. Considering that Miller et al [23] collected all the VRS studies, including training scenarios occurring outside hospitals and in-hospital events, the increase in VRS studies on in-hospital training programs is noticeable. In addition, it was observed that the number of studies published during the past 5 years (from 2016 to 2020) was greater than that published in the older but longer period between 2006 and 2015 (ie, 7 [29-35] vs 5 studies [6,24,26-28]).

The scale of participants is 1 of the important variables to draw conclusive findings from studies. There were 4 (33%) of 12 studies [6,31,33,35] that considered the scale of participants by conducting VRS training programs with more than 50

participants. This increased from 2 in the previous integrated review by Miller et al [23] and was double. In addition, all 4 studies [6,31,33,35] were conducted in the recent 5 years. These observations suggest that future studies would put more emphasis on this important variable and the potential outcomes can be more conclusive.

It was observed that 7 (58%) of 12 studies [29-35] since the past 5 years used consistent development platforms, such as the Virtual Electrosurgical Skill Trainer (VEST) and Unity, for the clinical implementation of VRS training programs. The use of established development platforms enables rapid implementation, easy scenario replication, and, therefore, increased feasibility in conducting VRS-based clinical training studies. This observation can somewhat explain the increasing trends in the number of VRS studies and studies with large-scale participants.

It was also observed that the details of evaluation processes were not fully described. There were only 3 (25%) of 12 studies [24,28,30] that explicitly stated the skill levels and experiences of health care participants, and 1 (8%) study [26] even lacked details of the participant number. Demographic details can be 1 of the influential variables to allow in-depth interpretation of outcomes from studies (eg, how learning effects work for different levels of expertise, as shown in the study by Pucher et al [28]). Similarly, qualitative analysis in the participants' experience was described using Liker-type surveys, postexperience interviews, and focus group open discussions, where there were little details on the reliability and validity of coding of qualitative data from interviews or focus groups. Such a lack was consistently found in the previous integrated review by Miller et al [23]. This integrated review, therefore, suggests that process details need to be sufficiently described in future studies, which will facilitate future investigators to draw deep insights into the value of VRS.

It was noted that the baseline performance comparisons of different studies highly varied: 3 (25%) of 12 studies compared VRS training programs among participants where some had previous disaster preparedness training experiences prior to study enrollment; 6 (50%) studies [6,26,27,30-32,34,35] compared VRS training effects against the experiences of live drills; and the remaining 3 (25%) studies [28,29,33] did not explicitly state the comparison approach. This may suggest that some efforts on developing the standard of VRS training program comparison protocols are meaningful, and this will allow for objective and quantitative comparison among different studies and draw general insights that are applicable to different disaster training programs.

Conclusion

The findings from this integrated review suggest that VRS could be a viable, cost-effective approach for health care professional training in in-hospital disaster preparedness. The reproducibility, just-in-time training, and repeatability features of VRS, along with its low cost of clinical implementation, suggest that VRS potentially represents a competitive adjunct to existing training approaches.

As VR continues to evolve in all technological aspects, it is anticipated that studies using VRS can become more vitalized in the clinical domain, while addressing currently unsolved issues. As an example of issues, studies with a massive number of participants with sophisticated assessment tools can be performed with more detailed and rigorous interventions and measurement of long-term retention.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT; no. 2020R1G1A1101300) and also by the Gachon University research fund of 2019 (GCU-2019-0776). The author gratefully acknowledges the contributions of SeungYoon Kim, a senior student of computer science at Gachon University, in the study search and initial screening.

Conflicts of Interest

None declared.

References

1. Federal Emergency Management Agency. Disaster Declarations by Year. URL: http://www.fema.gov/disasters/grid/year [accessed 2021-05-21]
2. Blau M. IRMA Forces At Least 35 Hospitals to Evacuate: Here's a Rundown. URL: https://www.statnews.com/2017/09/09/irma-hospital-evacuations-rundown/ [accessed 2021-05-21]
3. Labrague L, Hammad K, Gloe D, McEnroe-Petitte DM, Fronda D, Obeidat A, et al. Disaster preparedness among nurses: a systematic review of literature. Int Nurs Rev 2018 Mar 14;65(1):41-53. [doi: 10.1111/inr.12369] [Medline: 28295314]
4. Wikipedia. Memorial Medical Center and Hurricane Katrina. URL: https://en.wikipedia.org/wiki/Memorial_Medical_Center_and_Hurricane_Katrina [accessed 2021-10-24]
5. Surge Hospitals: Providing Safe Care in Emergencies. Joint Commission. URL: https://www.jointcommission.org/-/media/deprecated-unorganized/imported-assets/tjc/system-folders/topics-library/surge_hospitalpdf.pdf [accessed 2021-05-21]
6. Farra S, Smith S, Gillepsie G, Nicely S, Ulrich D, Hodgson E. Decontamination training with and without virtual reality simulation. Adv Emerg Nurs 2015 Apr;37(2):125-133. [doi: 10.1097/tme.0000000000000059]
7. Chapman K, Arbon P. Are nurses ready? Australas Emerg Nurs J 2008 Aug;11(3):135-144. [doi: 10.1016/j.aenj.2008.04.002]

8. Slepski L, Littleton-Kearney M. Disaster nursing educational competencies. In: Powers R, Daily E, editors. International Disaster Nursing. UK: Cambridge University Press; 2010:549-560.
9. Bhatti I, Jones K, Richardson L, Foreman D, Lund J, Tierney G. E-learning vs lecture: which is the best approach to surgical teaching? Colorectal Dis 2011 Apr;13(4):459-462. [doi: 10.1111/j.1463-1318.2009.02173.x] [Medline: 20041922]
10. Seth A, Vance JM, Oliver JH. Virtual reality for assembly methods prototyping: a review. Virtual Real 2010 Jan 22;15(1):5-20. [doi: 10.1007/s10055-009-0153-y]
11. Kavanagh S, Luxton-Reilly A, Wuensche B, Plimmer B. A systematic review of virtual reality in education. Themes Sci Technol Educ 2017;10(2):85-119.
12. Wang P, Wu P, Wang J, Chi H, Wang X. A critical review of the use of virtual reality in construction engineering education and training. Int J Environ Res Public Health 2018 Jun 08;15(6):1204 [FREE Full text] [doi: 10.3390/ijerph15061204] [Medline: 29890627]
13. Lanzo JA, Valentine A, Sohel F, Yapp AYT, Muparadzi KC, Abdelmalek M. A review of the uses of virtual reality in engineering education. Comput Appl Eng Educ 2020 Apr 22;28(3):748-763. [doi: 10.1002/cae.22243]
14. Pottle J. Virtual reality and the transformation of medical education. Future Healthc J 2019 Oct 11;6(3):181-185 [FREE Full text] [doi: 10.7861/fhj.2019-0036] [Medline: 31660522]
15. Lerner D, Mohr S, Schild J, Göring M, Luiz T. An immersive multi-user virtual reality for emergency simulation training: usability study. JMIR Serious Games 2020 Jul 31;8(3):e18822 [FREE Full text] [doi: 10.2196/18822] [Medline: 32735548]
16. Gulick V, Graves D, Ames S, Krishnamani PP. Effect of a virtual reality-enhanced exercise and education intervention on patient engagement and learning in cardiac rehabilitation: randomized controlled trial. J Med Internet Res 2021 Apr 15;23(4):e23882 [FREE Full text] [doi: 10.2196/23882] [Medline: 33856355]
17. Kamińska D, Sapiński T, Wiak S, Tikk T, Haamer R, Avots E, et al. Virtual reality and its applications in education: survey. Information 2019 Oct 16;10(10):318. [doi: 10.3390/info10100318]
18. Clus D, Larsen ME, Lemey C, Berrouiguet S. The use of virtual reality in patients with eating disorders: systematic review. J Med Internet Res 2018 Apr 27;20(4):e157 [FREE Full text] [doi: 10.2196/jmir.7898] [Medline: 29703715]
19. Perron JE, Coffey MJ, Lovell-Simons A, Dominguez L, King ME, Ooi CY. Resuscitating cardiopulmonary resuscitation training in a virtual reality: prospective interventional study. J Med Internet Res 2021 Jul 29;23(7):e22920 [FREE Full text] [doi: 10.2196/22920] [Medline: 34326040]
20. Nayyar A, Mahapatra B, Nhuong Le D, Suseendran G. IJET 2018 Apr 20;7(2.21):156. [doi: 10.14419/ijet.v7i2.21.11858]
21. Mehrfard A, Fotouhi J, Taylor G, Forster T, Armand M, Navab N, et al. Virtual reality technologies for clinical education: evaluation metrics and comparative analysis. Comput Methods Biomech Biomed Eng: Imaging Vis 2020 Oct 26;9(3):233-242. [doi: 10.1080/21681163.2020.1835559]
22. Ragazzoni L, Barco A, Echeverri L, Conti A, Linty M, Caviglia M, et al. Just-in-time training in a tertiary referral hospital during the COVID-19 pandemic in Italy. Acad Med 2021 Mar:336-339. [doi: 10.1097/acm.0000000000003575]
23. Farra SL, Miller ET. Integrative review: virtual disaster training. JNEP 2012 Nov 04;3(3):93-101. [doi: 10.5430/jnep.v3n3p93]
24. Heinrichs WL, Youngblood P, Harter P, Kusumoto L, Dev P. Training healthcare personnel for mass-casualty incidents in a virtual emergency department: VED II. Prehosp Disaster Med 2010 Jun 28;25(5):424-432. [doi: 10.1017/s1049023x00008505] [Medline: 21053190]
25. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
26. Roy MJ, Sticha DL, Kraus PL, Olsen DE. Simulation and virtual reality in medical education and therapy: a protocol. Cyberpsychol Behav 2006 Apr;9(2):245-247. [doi: 10.1089/cpb.2006.9.245] [Medline: 16640488]
27. LeRoy Heinrichs W, Youngblood P, Harter PM, Dev P. Simulation for team training and assessment: case studies of online training with virtual worlds. World J Surg 2008 Feb 11;32(2):161-170. [doi: 10.1007/s00268-007-9354-2] [Medline: 18188640]
28. Pucher P, Batrick N, Taylor D, Chaudery M, Cohen D, Darzi A. Virtual-world hospital simulation for real-world disaster response: design and validation of a virtual reality simulator for mass casualty incident management. J Trauma Acute Care Surg 2014 Aug;77(2):315-321. [doi: 10.1097/ta.0000000000000308]
29. Dorozhkin D, Olasky J, Jones DB, Schwaitzberg SD, Jones SB, Cao CGL, SAGES FUSE Committee. OR fire virtual training simulator: design and face validity. Surg Endosc 2017 Sep 30;31(9):3527-3533 [FREE Full text] [doi: 10.1007/s00464-016-5379-7] [Medline: 28039649]
30. Dubovsky SL, Antonius D, Ellis DG, Ceusters W, Sugarman RC, Roberts R, et al. A preliminary study of a novel emergency department nursing triage simulation for research applications. BMC Res Notes 2017 Jan 03;10(1):15-12 [FREE Full text] [doi: 10.1186/s13104-016-2337-3] [Medline: 28057048]
31. Farra S, Hodgson E, Miller ET, Timm N, Brady W, Gneuhs M, et al. Effects of virtual reality simulation on worker emergency evacuation of neonates. Disaster Med Public Health Prep 2019 Apr 8;13(2):301-308 [FREE Full text] [doi: 10.1017/dmp.2018.58] [Medline: 30293544]
32. Sankaranarayanan G, Wooley L, Hogg D, Dorozhkin D, Olasky J, Chauhan S, et al. Immersive virtual reality-based training improves response in a simulated operating room fire scenario. Surg Endosc 2018 Aug 25;32(8):3439-3449. [doi: 10.1007/s00464-018-6063-x] [Medline: 29372313]

33. Lovreglio R, Gonzalez V, Feng Z, Amor R, Spearpoint M, Thomas J, et al. Prototyping virtual reality serious games for building earthquake preparedness: the Auckland City Hospital case study. Adv Eng Inform 2018 Oct;38:670-682. [doi: 10.1016/j.aei.2018.08.018]
34. Rossler KL, Sankaranarayanan G, Duvall A. Acquisition of fire safety knowledge and skills with virtual reality simulation. Nurse Educator 2019;44(2):88-92. [doi: 10.1097/nne.0000000000000551]
35. Farra SL, Gneuhs M, Hodgson E, Kawosa B, Miller E, Simon A. Comparative cost of virtual reality training and live exercises for training hospital workers for evacuation. Comput Inform Nurs 2019 Sep;37(9):446-454. [doi: 10.1097/cin.0000000000000540]
36. Krueger WCF. The effect of overlearning on retention. J Exp Psychol 1929;12(1):71-78. [doi: 10.1037/h0072036]
37. Rahn R, Moraga R. The study of knowledge retention and increased learning through the use of performance based tasks. 2007 Presented at: IIE Annual Conference and Expo; 2007; Nashville URL: https://www.academia.edu/download/36255660/B13F_1_-_Copy.pdf

Abbreviations

CG: control group
EG: experiment group
EMB: Embase
MCI: mass casualty incident
PMD: PubMed
RCT: randomized controlled trial
VEST: Virtual Electrosurgical Skill Trainer
VR: virtual reality
VRS: virtual reality simulation
WOS: Web of Science

Edited by R Kukafka; submitted 25.05.21; peer-reviewed by F Ghezelbash, R Ciorap, R Lundin, A Joshi; comments to author 08.08.21; revised version received 21.11.21; accepted 17.12.21; published 28.01.22.

Please cite as:
Jung Y
Virtual Reality Simulation for Disaster Preparedness Training in Hospitals: Integrated Review
J Med Internet Res 2022;24(1):e30600
URL: https://www.jmir.org/2022/1/e30600
doi:10.2196/30600
PMID:35089144

Review

Best Practices and Lessons Learned for Action Research in eHealth Design and Implementation: Literature Review

Kira Oberschmidt[1,2], MSc; Christiane Grünloh[1,2], PhD; Femke Nijboer[2], PhD; Lex van Velsen[1,2], PhD

[1]eHealth Cluster, Roessingh Research and Development, Enschede, Netherlands

[2]Biomedical Signals and Systems Group, University of Twente, Enschede, Netherlands

Corresponding Author:
Kira Oberschmidt, MSc
eHealth Cluster
Roessingh Research and Development
PO Box 310
Enschede, 7500 AH
Netherlands
Phone: 31 88 0875 777
Email: k.oberschmidt@rrd.nl

Abstract

Background: Action research (AR) is an established research framework to introduce change in a community following a cyclical approach and involving stakeholders as coresearchers in the process. In recent years, it has also been used for eHealth development. However, little is known about the best practices and lessons learned from using AR for eHealth development.

Objective: This literature review aims to provide more knowledge on the best practices and lessons learned from eHealth AR studies. Additionally, an overview of the context in which AR eHealth studies take place is given.

Methods: A semisystematic review of 44 papers reporting on 40 different AR projects was conducted to identify the best practices and lessons learned in the research studies while accounting for the particular contextual setting and used AR approach.

Results: Recommendations include paying attention to the training of stakeholders' academic skills, as well as the various roles and tasks of action researchers. The studies also highlight the need for constant reflection and accessible dissemination suiting the target group.

Conclusions: This literature review identified room for improvements regarding communicating and specifying the particular AR definition and applied approach.

(J Med Internet Res 2022;24(1):e31795) doi:10.2196/31795

KEYWORDS

action research; eHealth; best practices; lessons learned

Introduction

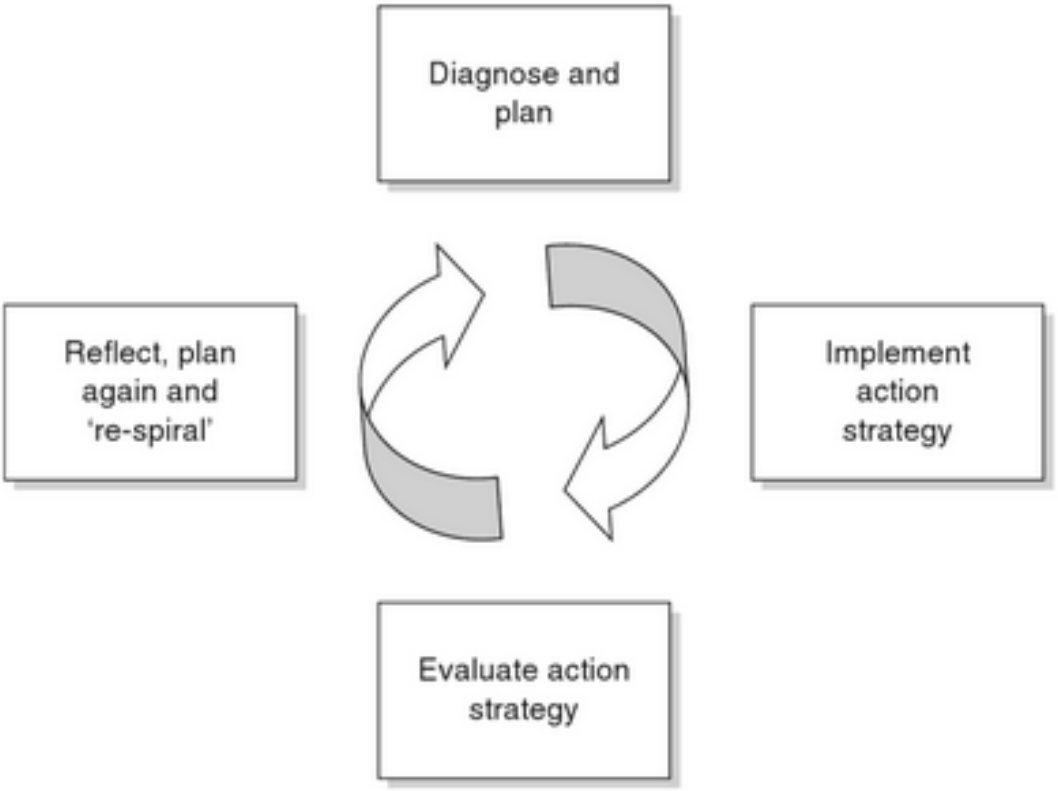
The way health care is organized and executed is of great societal concern, as it affects our quality of life. Hence, health care systems and eHealth technologies used to support health care should be designed in a way that meets the needs and expectations of their stakeholders. One way of doing this is through action research (AR). According to Bradbury and Lifvergren [1], AR in health care "seeks to (1) improve patient experiences and the health of populations, (2) reduce the per capita cost, (3) improve the work life of those who deliver care, and (4) bring health care providers into circumstances that allow for continuous learning together with patients." AR has been used as a research framework in nursing and health care, for example, to improve the quality of patient care and investigate changes in action [2]. AR is a collaborative approach, where people affected by the change envisioned in AR become active members of the research team. AR is often used in the design of eHealth systems. However, existing literature reviews of AR in eHealth predominantly focus on the development of new frameworks [3-5] but not on how eHealth AR is currently carried out. Therefore, this literature review outlines the state of the art of AR in eHealth design.

eHealth projects cover a wide variety of topics and technologies and can therefore greatly benefit patients, professionals, and many other health care stakeholders. However, to gain the most from eHealth systems and technologies, it is crucial that they match with what is needed in practice [6]. To ensure such a

match, Van Gemert-Pijnen and colleagues [6] suggest, among other things, working together with relevant stakeholders in all stages of the project, implementing the study results in practice, and continuously evaluating the process. Similarly, co-design has been mentioned as a useful technique for creating eHealth systems that suit the needs of the end users [3]. These ideas fit well with the principles of AR, which will be outlined below.

Definitions of AR have changed over the years. AR originated with Kurt Lewin [7], who described it as several consecutive circles of planning, action, and reflection. These cycles are shown in Figure 1, developed by Williamson and colleagues [2]. In later definitions, the cyclical nature of AR remains one of its key features. Reason and Bradbury [8], who build on Lewin's work, define AR as research that (1) involves stakeholders not only as participants but also as members of the research team, (2) consists of (at least) 1 cycle of planning, action, and reflection, (3) establishes direct changes, and (4) then evaluates those changes in and with the community. Their work [8] includes many interesting examples of AR from various fields. Furthermore, Bradbury and colleagues defined 7 "choice points for quality in action research" [9], criteria that can be used to plan, conduct, report, and assess AR projects.

Figure 1. Action research cycles (adapted from Kurt Lewin [7] by Williamson and colleagues [2]).



Within AR, different variations exist, such as action design research (ADR) or participatory action research (PAR). Usually, there is agreement on the main principles of AR explained earlier, but some authors or groups emphasize some aspects over others. For example, as the name suggests, ADR incorporates elements of design research into AR [10], whereas PAR highlights the involvement of the community [11]. For a more detailed overview of the similarities and differences between some of these approaches, see Williamson et al [2] or Coghlan and Brannick [11].

In general, AR and AR approaches such as ADR are similar to participatory design (PD) approaches that are used in human computer interaction (HCI) research. However, AR emphasizes reflection on and learning from the process that was carried out, whereas the main aim of PD is to create a solution [12]. AR, as opposed to PD, does not start with a clear goal of what needs to be developed but defines this throughout the process together with stakeholders. Additionally, AR is more immersive and calls for stakeholder involvement for a longer period of time due to its iterative cycles [13]. Nevertheless, in some cases, studies that are described as PD-related ones also meet Reason and Bradbury's criteria [8] for AR [14]. Hayes [12] argues that AR and HCI research can supplement each other, as both often provide solutions on a local scale. As Hughes [15] describes, there is no standard way of implementing AR in health care due to the broadness of the field. Instead, there is a variety regarding the why, how, and with whom AR in health care is carried out [15,16]. For example, levels of stakeholder engagement and the context in which AR takes place can vary [16]. Other differences among AR studies include the topic, country, project duration, main target group, and methods used. Therefore, these aspects are considered in this review. The purpose of this review is to give an overview of the current literature on eHealth AR and summarize the best practices and points of improvement for future eHealth AR projects. Special attention is paid to the contextual variables of the research (eg, setting, duration, number of stakeholders), as this is expected to influence the outcomes, best practices, and points of improvement of a study. To provide an overview of AR in eHealth, this literature review addresses the following subquestions:

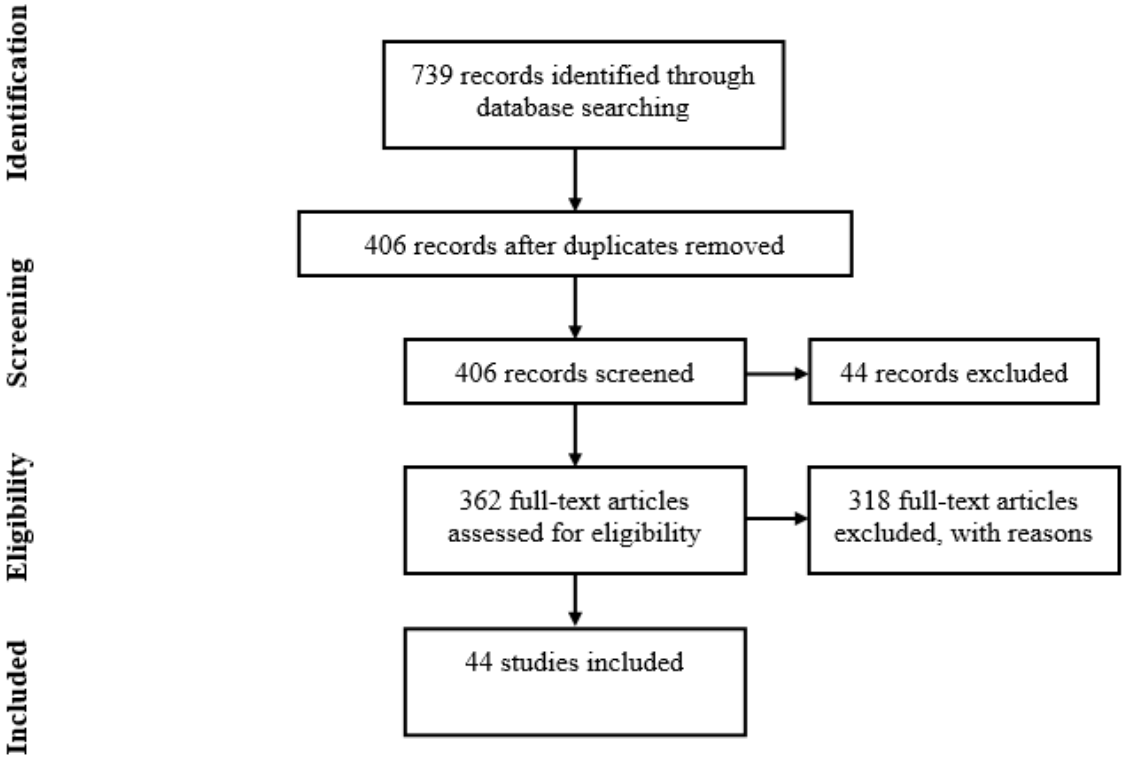
1. What is the context of AR eHealth projects?
2. How do eHealth AR studies define and operationalize AR?
3. What are the best practices for conducting AR in concrete eHealth studies?
4. What are the lessons learned from conducting AR in concrete eHealth studies?

Methods

Study Selection and Screening

The search was carried out in June 2020. PubMed, Scopus, and Google Scholar were searched using combinations of the search terms "action research" or "participatory design" and "eHealth," "health technology," "digital health," or "telemedicine." PubMed was chosen for its extensive medical database, and Scopus and Google Scholar were chosen as large scientific databases. Searching for "action research" turns up articles that include similar and related keywords like "participatory action research," "action design research," or "action-based research." "Participatory design" was included as a search term because PD has significant overlap with AR, and both are sometimes used to supplement each other. The list of synonyms for "eHealth," although not exhaustive, is expected to cover the various facets of the field. The initial search yielded 739 results. Articles were included if they (1) used and explicitly mentioned AR and (2) were about eHealth or health technology. Papers were excluded if they (1) were not written in English, (2) only included a study protocol but did not report results, or (3) only included a review of other articles. Full-text screening of the same 15 articles was performed by 2 authors (KO and CG); the authors discussed whether to include the studies until an agreement was reached. Next, the first author screened the full texts of the remaining articles, with some exceptions where a second opinion was necessary. These were again discussed between the first and second authors until an agreement was reached. Ultimately, 44 articles were included, reporting on 40 different projects. Figure 2 shows the flowchart of the inclusion process.

Figure 2. Inclusion flowchart of the literature search and screening process.



Data Extraction

For each study, the definition of AR that was provided by the authors, and the related AR approaches that they cited (if any) were extracted. Additionally, information about contextual variables of the study was derived. Specifically, we identified the topic, country, organizational context, project duration, types of stakeholders involved, the main target group of the research, and methods used. The types of involved stakeholders were grouped according to the framework described by Schiller et al [17], in which they define the main stakeholder categories as the public, policy makers, and governments, the research community, practitioners and professionals, health and social service providers, civil society organizations, and private businesses. Finally, the best practices and lessons learned were derived. The best practices and lessons learned were activities that could move forward and benefit the AR project, without necessarily being recognized as standard components of AR. The difference between what was seen as a best practice and as a lesson learned was based on the timing and reporting of these actions. An activity was labeled as a best practice if researchers already planned their project with this in mind (eg, mentioning it in the description of the methods). On the other hand, lessons learned were those points that researchers came to know during their project, which were reported mainly in the discussion section. From the first 5 articles, the best practices and lessons learned were extracted by 2 authors (KO and CG), and they

compared their results. The remaining data were extracted by 1 author (KO) in consultation with the second author where a second opinion was necessary. Furthermore, 5 authors published not 1 but 2 papers about their project. For these papers, the same study context was described whereas the definition of and approach to AR and the best practices and lessons learned were reported separately, as these sometimes differed between the articles. A reflection on 2 projects was included in 1 article. In this case, each project context was reported separately whereas only 1 AR definition and approach as well as one set of best practices and lessons learned were outlined.

Synthesis

A general overview of all the included studies describing the AR approach, AR definition, and contextual variables was obtained. The contextual variables (topic, location, target group, stakeholders, duration, and methods used) were categorized. Furthermore, the studies were mapped in a matrix based on the study duration and the types and number of different stakeholders that participated in the study. The contextual data were coded and categorized inductively. To identify which AR approach was the most used, the citation frequency of each approach in the included studies was recorded. Furthermore, the cited AR approaches that were available were accessed and checked for cross-referencing. All cited AR definitions were mapped to show the relationship between them. The AR definitions used, best practices, and lessons learned were coded by 1 author (KO). The best practices and lessons learned were coded individually first and then combined for both categories.

Results

Context

The setting of the included studies was described based on 6 categories (topic, location, duration, involved stakeholders, target group, and methods). Multimedia Appendix 1 presents all the categories and the description of the setting for each study. The most common aspects of each category will be discussed below.

Topic

We identified 9 broader categories of the research topics in the 44 included studies (see Table 2.1 in Multimedia Appendix 2 for the full list). The most common were home care and telemonitoring, and health promotion and education (both n=8), followed by electronic medical records and health information systems (n=7), and mental health services (n=5).

Location

The studies were set in 21 different countries, Australia being the most common (n=5) followed by the United States (n=4) and Canada, Sweden, and the United Kingdom (all n=3). Some studies from nonwestern countries, like Tanzania or Colombia were included, but no country was represented more than once or twice. Within the different countries, studies took place in various contexts, the most prevalent of which were rural areas (n=6) and hospitals (n=5). All contexts and countries can be found in Tables 2.2 and 2.3 of Multimedia Appendix 2.

Target Groups

Among the 44 studies, 2 studies explicitly focused on 2 different target groups at the same time, whereas all other studies had 1 main target group. In most cases, the target groups were patients (n=11). Of these, the most common group was patients with cancer (n=3). There were 6 studies each focusing on clinicians as well as children and young adults, and 5 studies targeted older adults (see Table 2.4 in Multimedia Appendix 2 for the full list of target groups).

Stakeholders

In many cases, several stakeholders were included in the study, up to 6 different types of stakeholders included in some cases. In summary, 20 different types of stakeholders were involved (see Table 2.5 in Multimedia Appendix 2 for the full list). Health care workers (n=18) and patients and their representatives (n=12) were involved the most, followed by governmental bodies (n=9) and general nonmedical staff members (n=8). When clustering these stakeholder types according to the framework defined by Schiller and colleagues [17], the largest group consisted of practitioners and professionals (n=48), followed by members of the public (n=38). Policy makers and government bodies (n=13), the research community (n=10), private businesses (n=6), and civil society organizations (n=3) were represented less often. The only group that was not represented at all included health and social service providers.

Duration

Not all of the 44 studies reported the duration of the project (n=7). Studies that did report the duration (n=33) lasted from a few months (n=5) to more than 10 years (n=2). The majority (n=13) of these studies reported a project duration between 2 and 3 years, and the average project duration was 2.7 years. Figure 3 shows the distribution of the 10 most frequently involved types of stakeholders for the different project durations in the 33 projects that reported the project duration. Stakeholder types are shown in the order of how many times they were involved in total; however, because some studies did not report project durations, the numbers in this graph differ from those described above. The 2 biggest stakeholder groups, health care workers and patients, were rarely, or in the case of patients even not at all, involved in long-term studies.

In Figure 4, the study duration is mapped against the number of different stakeholders that were involved in each of the 33 projects that reported a project duration. Studies that did not report the overall project duration are not included in the figure. Most of the included studies lasted for up to 2 years, including 2 or 3 stakeholder groups. There are some longer studies including more stakeholder groups.

Figure 3. Heat map showing the most commonly involved types of stakeholders against the project duration.

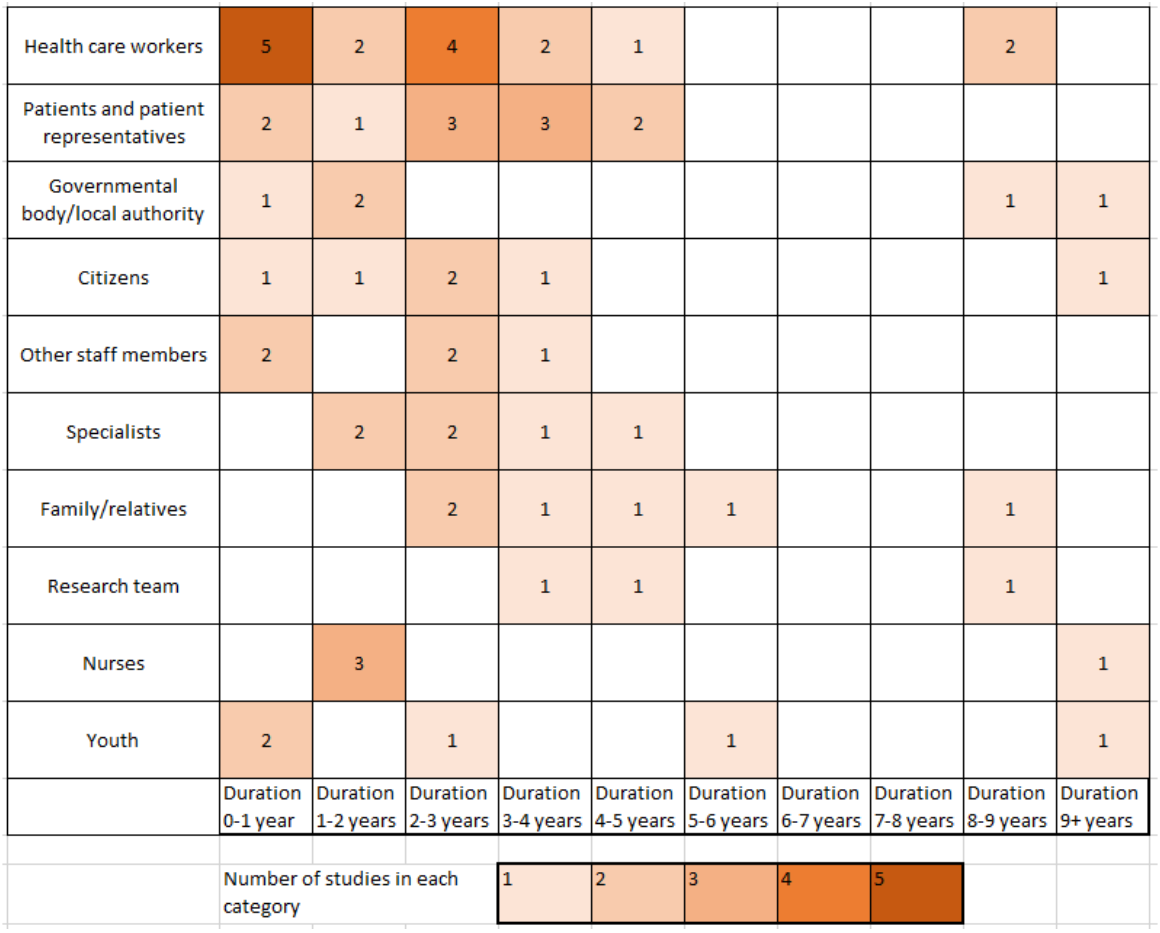
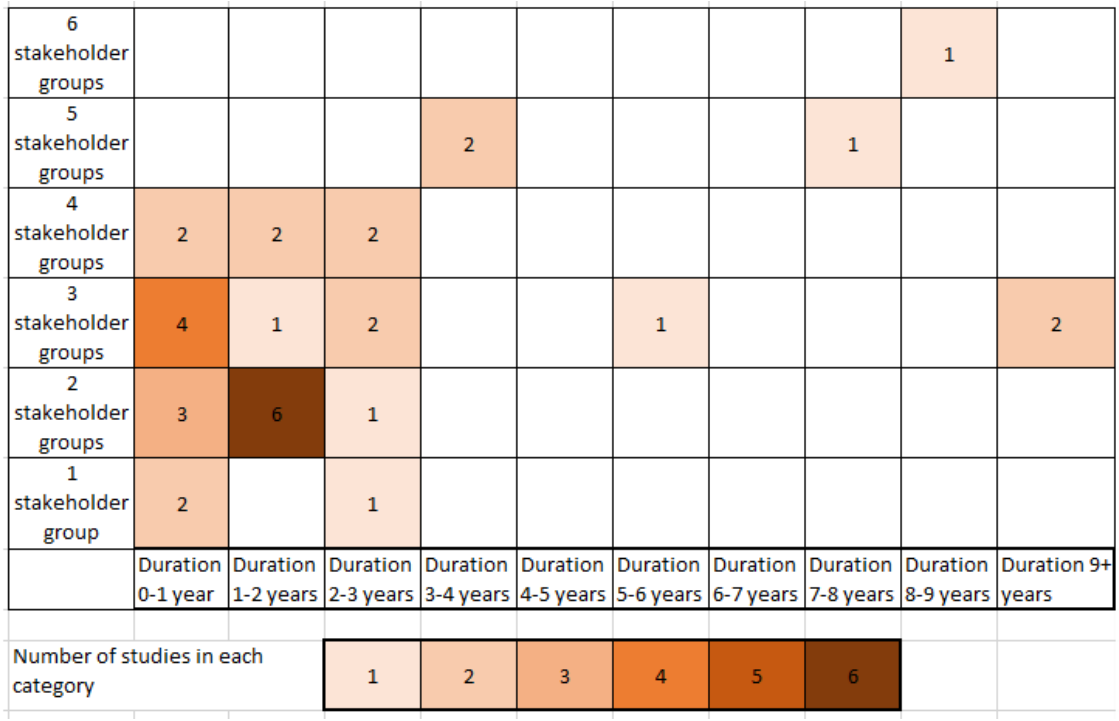


Figure 4. Heat map showing the number of stakeholders involved against the project duration.



Research Methods Used

As mentioned earlier, AR is a framework that does not advise the use of a single methodology, and studies can therefore include a variety of different research methods. Most of the 44 included studies indeed used several methods, with some studies employing up to 6 different methods. Interviews were used most frequently (n=24), followed by focus groups (n=22), workshops

(n=14), and surveys (n=13). On average, studies used nearly 3 different methods (average 2.8). All methods can be found in Table 2.6 of Multimedia Appendix 2.

AR Definitions

The articles contained 44 definitions of AR. They could be grouped according to 4 different aspects that they emphasized. First, 21 studies emphasized that in AR projects, practitioners and other stakeholders become (co)researchers (n=21). Second, AR is a cyclical process that includes different stages (n=19). Third, 14 studies described how AR focuses on solving a practical issue and aims to extend research knowledge. The fourth aspect was that AR takes place in a community setting (n=10). Further, 2 studies included 3 of these aspects in their definitions, and only 2 other studies mentioned all 4 aspects. Most studies included either 1 (n=16) or 2 (n=17) of the aspects, whereas 7 studies included none of these points in their definition or did not at define AR in detail. Table 1 provides an overview of the number of mentions per aspect and the studies mentioning these aspects.

Table 1. Number of mentions and studies mentioning the aspects of the AR definition.

Aspect of the AR definition	Number of articles that define AR including this aspect, n (N=44)	References
Practitioners and other stakeholders being (co)researchers	21	[18-38]
Cyclical process including different stages	19	[23,24,26,27,30,31,34,35,38-48]
Aiming to solve a practical problem and extend academic knowledge	14	[20,21,25,30,32,36-39,42,43,49-51]
Research taking place in a community setting	10	[19,20,23,28-30,38,52-54]

AR Approaches

Table 2 gives an overview of the AR approaches that were cited at least twice in the included articles. The AR approach was not cited in 4 studies. In some cases, different papers from the same authors were cited; however, as these eventually described the same approach, the citation count was added up. The most commonly cited approach was that proposed by Reason and Bradbury [8]. As described earlier, the key elements of this approach are that AR (1) involves stakeholders as coresearchers, (2) consists of plan, act, and reflect cycles, (3) makes a change in practice, and (4) evaluates the said changes in and with the community. Overall, most definitions share these main aspects but differ in terms of the aspects that are particularly emphasized. For example, Baskerville and colleagues [55] highlight the duality of practical work and scientific knowledge, whereas Baum and colleagues [56] underline the need for reflective practice that includes all stakeholders. Figures 5 and 6 depict the cited approaches in more detail. There are 3 independent researchers or groups that are mentioned as being the origin of AR, namely Lewin [7], Trist and colleagues [57], and Freire [58]. Wherever the origin of AR was mentioned, some cases have named 2 of these, as observed in Figure 5. The cited AR approaches also frequently refer to each other and sometimes authors collaborate with each other, for example on books about AR (see Figure 6). There are no very distinct groups conducting their own AR, but the different AR groups are often connected and build upon each other's work.

Table 2. Overview of the most cited action research approaches in the included articles per author or research group, including the number of citations.

Author(s)	Number of author citations	References	Action research approach paper(s) describing these approaches
Peter Reason and Hilary Bradbury	8	[23,28,35,37,42,44]	[8,59,60]
Robert N. Rapoport	4	[35-37,39]	[61]
David Avison and colleagues	3	[37,39,49]	[62]
Richard L. Baskerville and colleagues	3	[29,38,43]	[55,63,64]
Jørn Braa, Eric Monteiro, and Sundeep Sahay	3	[18,53,54]	[61]
Stephen Kemmis and Robin McTaggart	3	[19,44,51]	[66,67]
Fran Baum, Colin MacDougall, and Danielle Smith	2	[24,52]	[56]
Bob Dick and colleagues	2	[35,44]	[68,69]
Max Elden and Morten Levin	2	[32,52]	[70]
Colin Robson	2	[20,38]	[71]
Harvey A. Skinner, Oonagh Maley, and Cameron D. Norman	2	[45,72]	[72,73]
Gerald I. Susman and Roger D. Evered	2	[38,43]	[74]
Elizabeth Hart	2	[33]	[75,76]
Gillian R. Hayes	2	[47,77]	[12]

Figure 5. Overview of the action research approaches referred to in the included articles, indicating those papers that are mentioned as "the origin" of action research. Studies that either name an approach as being the origin of action research, or are being named as such, are highlighted in blue for better readability.

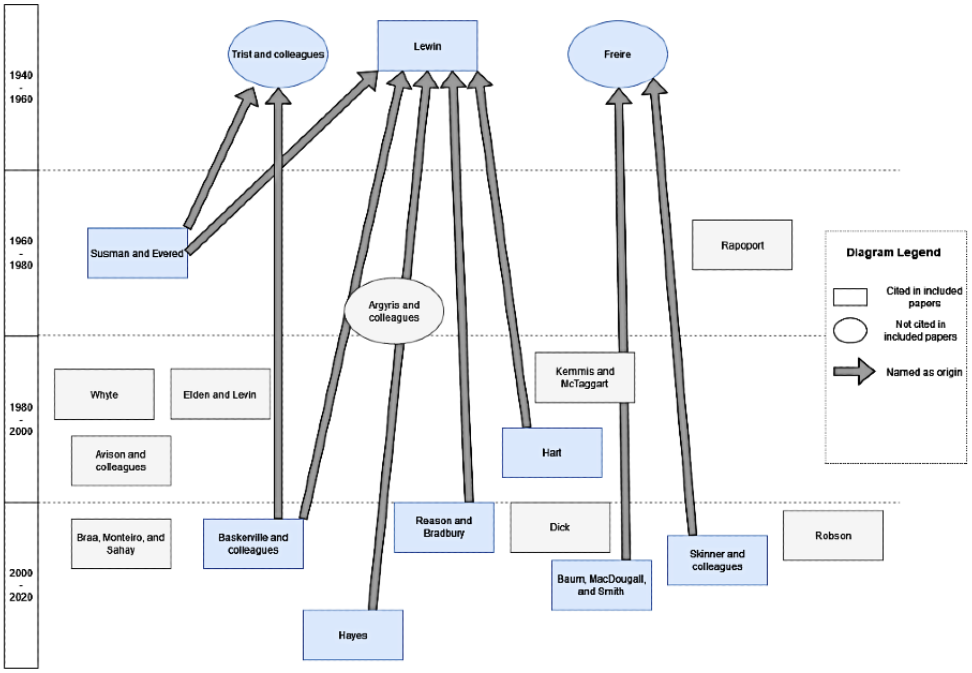
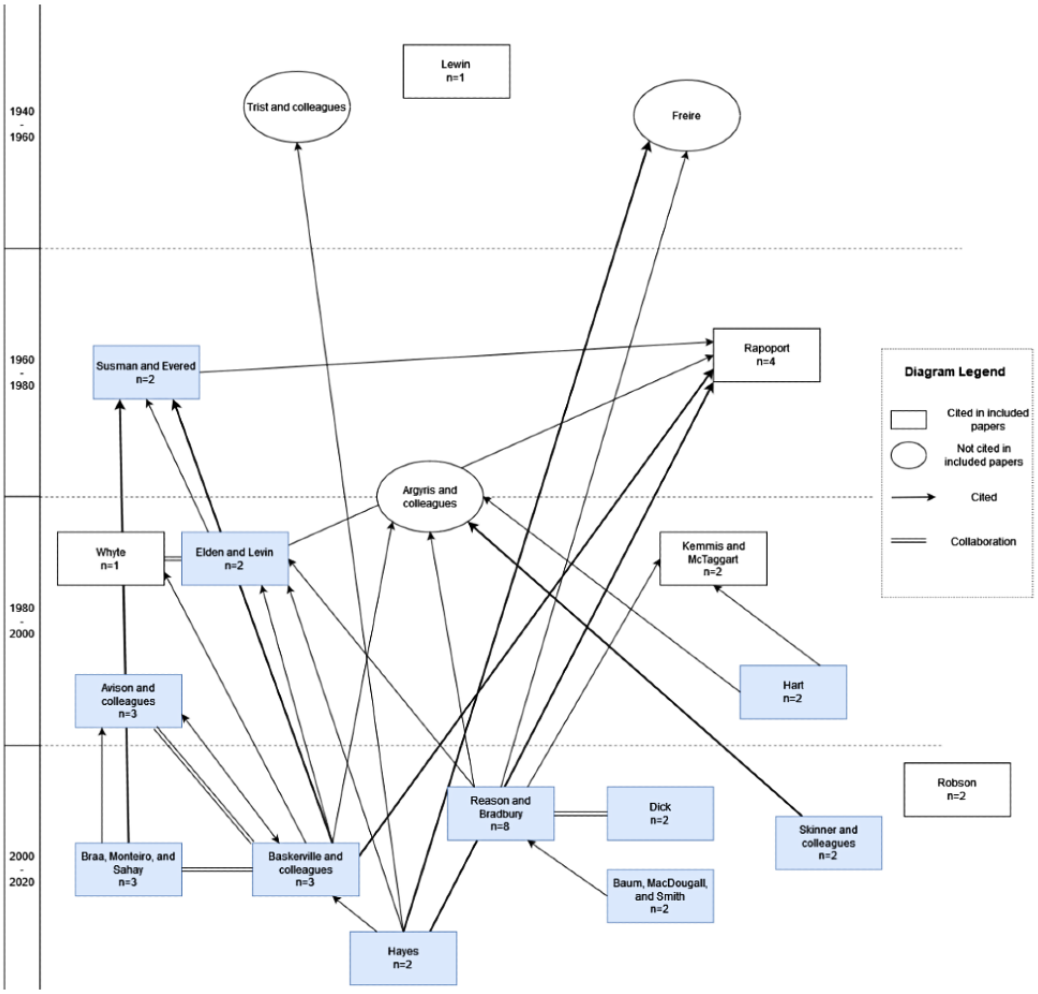


Figure 6. Overview of action research approaches referred to in the included articles. Arrows indicate citations between the action research approach papers. The number of times that the articles included in this review cited each approach is indicated in the box. We have used different arrow thicknesses for better readability. Blue boxes indicate those papers that were available and checked for citations.



Best Practices and Lessons Learned

As previously described, an activity was identified as a "best practice" if researchers already planned their project with this in mind (eg, mentioning it in the description of methods). Lessons learned were those points that researchers came to know during their project. These were mostly reported in the discussion section. In total, 85 best practices and 66 lessons learned were identified, which were clustered into 22 categories of best practices and 16 categories of lessons learned. Among the 44 papers, 3 papers did not indicate any best practices that they followed, whereas 12 papers did not include any identifiable lessons learned. There were 8 overlapping categories, identified as best practices in some articles and as lessons learned in others. These will be discussed in more detail below.

Best Practices

The identified best practices in the 44 studies were most often related to the use of a specific method (n=9), namely personas (n=2), world café, journey mapping, role play, scenarios, case studies, design cards, and mixing different types of data collection methods (all n=1). Other best practices were a continuous evaluation of the project and a reflection on the process by the research team (n=8). The importance of establishing active contact between researchers and stakeholders and raising the confidence and skills of stakeholders was emphasized by 7 studies. The improvement of stakeholder skills mainly referred to research and analytical skills, allowing stakeholders to set up their own studies or continue the work after the project was finished. There were several specific suggestions to improve the regular project team meetings, for example, to always use the same agenda or to share a common area (office space) to make contact easier. Some other best practices concern the reporting and presentation of outcomes (n=6). The complete list of best practice categories can be found in Table 3.

Table 3. Overview of all best practice categories and number of mentions per category (N=44).

Best practices category	Number of mentions, n
Process	
Recommends specific method	
Personas	2
World Café	1
Journey mapping	1
Role play	1
Scenarios	1
Case study	1
Design cards	1
Abstract vs personal methods of data collection	1
Continuous evaluation and reflection	8
Report or present results	
Share resources and findings (on the internet) allowing others to benefit from it	4
Present findings to the community or target group in a suitable manner	2
Start with close examination of context (observation and literature)	5
Agile development and Scrum	3
Combining action research with randomized controlled trials (RCTs)	2
Combining these 2 approaches	1
Keeping the line between stakeholders and researchers blurred and not performing RCTs	1
Gradual scaling up	2
Immediately resolve problems and apply lessons learned	2
Stakeholders and relationships	
Frequent or regular (face-to-face) meetings, active contact (eg, shared space), and same transparent agenda	7
Raising stakeholder confidence and skills (eg, analytical skills so that they can set up their own studies)	7
Clearly defining the role of each partner (equal involvement is not always good)	5
Finding committed stakeholders with intrinsic motivation (to carry on with the project after the researchers have left)	5
Reference group (with technical, juridical, and clinical expertise)	4
Stepping into each other's shoes (experiencing the other's tasks and familiarizing oneself with what the other does)	3
Investing in relationship between partners (also nonwork activities)	3
Adapting methods or schedules to the needs of stakeholders	3
Neutral position of the researcher (no steering or predetermined outcomes, serving as a communication link instead)	3
Patient- and stakeholder-generated content (eg, personas)	2
Different disciplines	2
Context and environment	
Living labs as context for action research	2
Actively encouraging pilot participation	2
Paying attention to economic or business values	3

Lessons Learned

Apart from the best practices, the lessons learned from each study were identified. The most common lessons learned were increasing stakeholder knowledge and skills (n=8) and continuous evaluation of the project and reflection on the process (n=6). Both of these had been identified as best practices in other articles (more on this overlap below). Recommendations for the use of specific methods were also common (n=5). Lessons learned regarding reporting, adapting the project to fit the needs of stakeholders, fostering a welcoming environment, and the questionable replicability of the research were each mentioned 4 times. All lessons learned are shown in Table 4.

Table 4. Overview of all lessons learned categories and number of mentions per category (N=44).

Lesson learned category	Number of mentions, n
Process	
Continuous reframing or renegotiation (flexibility), baby steps	6
Recommend specific method	
Field work	1
Randomized controlled trial	1
Case study	1
Action circles	1
Fun methods (quiz, game, puzzle) as learning opportunities	1
Reporting	
Open source	2
Higher level of sophistication necessary	1
Also include nonproject target group	1
Integration of literature	3
Regular meetings to check on progress and motivate the stakeholders (reality check)	2
Triangulation of data to decrease biases	2
End of an action research project	
Accompanying stakeholders until they find that the process is done	1
Action research leading to other collaborative activities	1
Commitment to action research necessary (eg, through specific funding)	1
Ethical restrictions	1
Immediate reflection impossible	1
Stakeholders and relationships	
Raising stakeholder confidence and skills, knowledge sharing	8
Tailoring to the needs of stakeholders	
Including action research in work schedule	1
Researchers taking over some of the stakeholders' usual tasks to make schedule less busy	1
Adequate feedback methods	1
Identifying unique strengths	1
Investing in relationship between partners	3
Accepting that participation is different for everyone and can change over time	3
Communication	
Language barrier	1
Finding a common language	1
Enthusiastic local "champion" to start the project and help keep people motivated	2
Involving authorities or local government (address issues at multiple levels)	2
Actively breaking down power structure	1
Context and environment	
Fostering a positive, welcoming environment for change	4
Questionable replicability	4
Active researcher involvement and presence in environment	2
Drawing attention to external influences	1
Ethical issues	1

Lesson learned category	Number of mentions, n
Diffusion of innovation	1
Organizational expectations	1

Overlapping Best Practices and Lessons Learned

As stated earlier, some aspects were identified as best practices in some articles and as lessons learned in others. In total, we identified 7 such overlapping aspects. Overall, the most mentioned aspect was the importance of raising stakeholder skills and confidence (n=15, where best practices= 7 and lessons learned=8). Many articles reported the need for stakeholders to learn new skills, for example related to academic research, or the need to be convinced about their ability to perform these tasks. Almost all the studies that reported this as a best practice or lesson learned involved health care professionals as stakeholders. Other commonly mentioned points were recommendations for specific methods, even though the suggested methods differed (n=14, where best practices=9 and lessons learned=5) and there was continuous reframing and evaluation of the project (n=14, where best practices=8 and lessons learned=6). Continuous reframing often referred to the iterations of planning, action, and evaluation in AR projects. Studies that described this mostly did not include this cyclical nature of AR in their definition of it. In total, there were 10 recommendations regarding the reporting and presentation of results (best practices=6 and lessons learned=4), for example calling for open and accessible publishing of outcomes. The best practices and lessons learned included recommendations about meeting regularly (n=9, where best practices=7 and lessons learned=2), adapting to the needs of stakeholders (n=8, where best practices=3 and lessons learned=5), and investing in the relationship between partners (n=6, where best practices=3 and lessons learned=3).

Chronology of Overlapping Best Practices and Lessons Learned

When observing the publication timeline, most of the overlapping aspects appeared as a lesson learned in earlier publications, and then as a best practice in papers published at a later point in time. This was the case regarding stakeholder skills, appearing as a lesson learned in 1999 [33] and as a best practice in 2016 [25]; continuous reframing of the project was a lesson learned in 2003 [19] and best practice in 2009 [42]; further, having regular meetings was a lesson learned in 2006 [72] and a best practice in 2018 [27], and adapting the research to stakeholder needs was a lesson learned in 2007 [32] and a best practice in 2016 [77]. Such a clear timeline could not be seen for accessible reporting, appearing as a lesson learned in 2017 [78] and a best practice in 2007 [45], and the relationship between partners appearing as a lesson learned in 2017 [36] and as a best practice in 2008 [38].

Discussion

Principal Results

To identify recommendations on how to conduct AR in eHealth studies, this literature review analyzed the setting, AR description, and best practices and lessons learned in 44 studies. The most important recommendations from this review, which will be discussed in more detail below, are as follows: actively raising stakeholder skills and confidence; fulfilling multiple roles and tasks as a researcher; fostering constant reflection and evaluation; ensuring open and accessible dissemination; reporting in a more structured and comprehensive way.

These recommendations are not exclusively related to eHealth, despite them being derived from a review of eHealth AR studies. Hence, it is possible that the recommendations are also relevant for AR in various other fields. Therefore, where possible, examples from different disciplines are discussed below to explain or supplement a recommendation.

Stakeholder Skills and Confidence

Being involved in a project as coresearcher can potentially increase stakeholders' confidence, besides teaching them new skills [79]. However, this does not happen automatically. Similar to our findings, the narrative review conducted by Harrison and colleagues [80] also identified educating the research team as the most important task when stakeholders are involved in health care research. Nevertheless, there is limited research on how skill training for stakeholders could look like, and this can vary greatly between studies. Stakeholders in some eHealth studies might need to learn content-related information [81], whereas other studies require methodological or statistical skills [54]. Researchers should provide adequate training and material for their project and encourage stakeholders to make use of it. The studies included in this review that recommended stakeholder skill training almost exclusively worked with health care professionals. The relationship between recommending skill training and working mainly with health care professionals remains unclear. A possible explanation could be that other stakeholder groups in other studies already had the necessary skills and thus did not require any additional training. Another possibility is that other stakeholders were not given the same roles that health care professionals held, and therefore, they did not need skill training. Finally, as we will discuss later, reporting of AR activities was not always very extensive. Thus, stakeholders outside the health care sector were possibly trained, and these studies did not report on this aspect. Generally, not all participants prefer the same level of engagement in a project, and researchers should respect these preferences [82].

Tasks and Roles of the Researcher

Different aspects of the role and tasks of the researcher in an AR project are discussed. Brydon-Miller and Aragón describe the many different tasks that action researchers need to fulfil as their "500 hats" [83]. These are not specific to eHealth studies, but they can occur in any AR study. As researchers and stakeholders have many varied duties, their roles are not fixed and might change over the course of the project [19]. One main task of the researchers that continues throughout the project is the need to foster a welcoming environment for all stakeholders

[42]. Researchers should also be present and actively involve themselves at a higher level than that needed in non-AR projects [38]. Additional AR-specific tasks for the researchers include investing in partner relationships [35] or breaking down power structures [28]. Generally, AR studies demand more self-reflection and awareness from the researchers than other projects and researchers should keep this in mind when entering an AR project.

Constant Reflection

The importance of continuous reframing and evaluation of the project was emphasized in several studies. Although evaluation is 1 of the AR cycles, studies providing recommendations on this topic rarely included this in their definition of AR. Owing to the lack of reports on AR cycles, which will be discussed below, it is unclear if these studies still followed the AR cycles without reporting on them. However, sometimes, it seems that periodic planned evaluation is not enough. Instead, the participants need to regularly reflect on the current status of the project and their role in it. Therefore, new AR projects should create suitable spaces for evaluation and reflection in ways that fit the projects and stakeholders. This is especially important because reflection can become difficult once a person is in the middle of the project [49]. Holeman and Kane [53] emphasize that reflection should not only take place within the project, but it should also be explicitly reported to help other researchers. If action researchers take reflection seriously and include honest evaluations in their publishing, the AR community members can learn from each other. Additionally, researchers and other stakeholders within the project learn and benefit from constant reflection [9].

Accessible Dissemination

Another important aspect concerns paying attention to open and understandable dissemination of results within the community and among researchers. Action researchers need to communicate findings to the academic world while also finding ways to inform the target group about the project in ways that suit the target users' needs. An example of open and accessible dissemination can be found in Canto-Farachala and Larrea [83]. They present the results of their AR project regarding territorial development on an interactive website, allowing others to learn from their work. However, it seems that accessible reporting is still not the norm in AR, as Avison and colleagues [62] describe that many AR studies are generally "published in books rather than as articles. Action researchers have large and complicated stories to tell." Future AR projects should attempt to narrate their stories in such a way that others can learn from them.

Comprehensive Reporting

The different way of describing AR studies also leads to another issue, incomplete and elusive reporting. Although most studies did provide at least a short description of what they saw as AR, 7 studies provided no definition at all. Additionally, there were only 4 studies that included 3 or all of the 4 aspects of the AR definition in their description. Even the most mentioned aspects appeared in less than half of the included papers. Even though most papers did cite an AR approach of definition, some did not. In combination with the often-limited descriptions of AR, this makes it difficult to obtain a clear picture of how AR is perceived and performed in a particular study. This resonates with what Bradbury and colleagues [9] describe as 1 of the quality points of AR, namely "action research process and related methods (should be) clearly articulated and illustrated." The best practices and lessons learned that were extracted from the included studies were seldom mentioned explicitly. Best practices were often hidden in the description of the project, without much reasoning. Similarly, lessons learned were often described as adaptations made during the project or as plans for the future. Although we observed that some lessons learned turned into best practices over time, we think that researchers could benefit more from each other's work by providing concrete recommendations. This review is a step in that direction. Both aspects show that the reporting of AR studies in eHealth can be improved to show more clearly what eHealth AR projects can look like and help others in setting up such projects with specific recommendations.

Limitations

Approximately a third of the included papers (14 out of 44) were published more than 10 years ago. This also means that some of the technologies that are described in the older papers are now relatively old. However, this literature review focuses mainly on the AR methodology and lessons learned about doing action research. Therefore, there was no exclusion criterium regarding the publication date of the papers.

The search yielded several PD-related papers. These papers could have been included, given that some definitions of PD are very similar to AR. However, as our aim was to provide an overview of how AR is done, these were excluded as the researchers of these studies themselves did not identify their studies as being related to AR (ie, not referring to, mentioning, or describing AR). Although this offers a clearer picture of how researchers conduct AR, it also creates a potential limitation in that best practices and lessons learned could be enriched from PD literature.

This overview of AR approaches focuses mostly on the interconnectedness among the approaches, without a comprehensive comparison of the content. Comparing the approaches with regard to the specific aspects of AR that they describe would be a review in and of itself, going beyond the scope of this current review. Therefore, we decided to focus on the definitions that the authors themselves provided even when they also cited AR approaches, as these are most likely to reflect their own vision of AR.

Conclusions

This review illustrates how AR is conducted in eHealth studies. Studies that fulfilled the inclusion criteria mainly took place in western countries and lasted for 2 to 3 years. Different stakeholders were involved, but the most commonly involved groups were health care professionals and patients. As for the methods used, most studies opted for focus groups and interviews. Even though many studies cited the AR approach proposed by Reason and Bradbury [8], their own definitions of AR were often not explicit in terms of how they implemented AR. Future projects should report their AR definition as well

as the best practices and lessons learned more clearly. Other recommendations include paying attention toward developing the skill and confidence of the stakeholders, being aware of the changing role of the researcher, frequently evaluating the project, and disseminating results in an understandable manner.

Authors' Contributions

KO performed the literature search and analysis and was a major contributor in designing the study and writing the manuscript. CG contributed to the design of the study, assisted with the search and analysis, and made major contributions to the manuscript. FN and LvV contributed to the design of the study and substantially revised the manuscript. All authors read and approved the final manuscript. This project has received funding from the European Union's Horizon 2020 research and innovation program (grant 857188).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full overview of all categories and description of settings for each category.
[XLSX File (Microsoft Excel File), 18 KB - jmir_v24i1e31795_app1.xlsx]

Multimedia Appendix 2
Full list of categories per setting variable.
[XLSX File (Microsoft Excel File), 16 KB - jmir_v24i1e31795_app2.xlsx]

References

1. Bradbury H, Lifvergren S. Action research healthcare: focus on patients, improve quality, drive down costs. Healthc Manage Forum 2016 Nov;29(6):269-274. [doi: 10.1177/0840470416658905] [Medline: 27770047]
2. Williamson GR, Bellman LC, Webster J. Action Research in Nursing and Healthcare. London: Sage; 2011.
3. Eyles H, Jull A, Dobson R, Firestone R, Whittaker R, Te Morenga L, et al. Co-design of mHealth delivered interventions: a systematic review to assess key methods and processes. Curr Nutr Rep 2016 Jul;5(3):160-167. [doi: 10.1007/s13668-016-0165-7]
4. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
5. Moore G, Wilding H, Gray K, Castle D. Participatory methods to engage health service users in the development of electronic health resources: systematic review. J Particip Med 2019 Feb;11(1):e11474 [FREE Full text] [doi: 10.2196/11474] [Medline: 33055069]
6. van Gemert-Pijnen JE, Nijland N, van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A holistic framework to improve the uptake and impact of eHealth technologies. J Med Internet Res 2011 Dec;13(4):e111 [FREE Full text] [doi: 10.2196/jmir.1672] [Medline: 22155738]
7. Lewin K. Action research and minority problems. Journal Soc Issues 1946;2(4):34-46. [doi: 10.1111/j.1540-4560.1946.tb02295.x]
8. Reason P, Bradbury H. The Sage Handbook of Action Research: Participative Inquiry and Practice. London: Sage; Jul 2013.
9. Bradbury H, Glenzer K, Ku B, Columbia D, Kjellström S, Aragón A, et al. What is good action research: quality choice points with a refreshed urgency. Action Res 2019 Mar;17(1):14-18. [doi: 10.1177/1476750319835607]
10. Sein MK, Henfridsson O, Purao S, Rossi M, Lindgren R. Action design research. MIS Quarterly 2011;35(1):37. [doi: 10.2307/23043488]
11. Coghlan D, Brannick T. Doing Action Research In Your Own Organization. London: Sage; 2005:1608-1616.
12. Hayes GR. The relationship of action research to human-computer interaction. ACM Trans Comput Hum Interact 2011 Jul;18(3):1-20. [doi: 10.1145/1993060.1993065]
13. Foth M, Axup J. Participatory design and action research: Identical twins or synergetic pair? In: Expanding Boundaries in Design: Proceedings Ninth Participatory Design Conference (Volume 2). 2006 Presented at: Ninth Participatory Design Conference 2006; August 1-5, 2006; Trento p. 93-96. [doi: 10.1145/1147261]
14. Clemensen J, Rothmann MJ, Smith AC, Caffery LJ, Danbjorg DB. Participatory design methods in telemedicine research. J Telemed Telecare 2017 Oct;23(9):780-785. [doi: 10.1177/1357633X16686747] [Medline: 28027678]
15. Hughes I. Action research in healthcare. In: Sage Handbook of Action Research. London: Sage; 2008.
16. Cordeiro L, Soares CB. Action research in the healthcare field: a scoping review. JBI Database System Rev Implement Rep 2018 Apr;16(4):1003-1047. [doi: 10.11124/JBISRIR-2016-003200] [Medline: 29634517]

17. Schiller C, Winters M, Hanson HM, Ashe MC. A framework for stakeholder identification in concept mapping and health research: a novel process and its application to older adult mobility and the built environment. BMC Public Health 2013 May;13(1):1-9 [FREE Full text] [doi: 10.1186/1471-2458-13-428] [Medline: 23639179]
18. Asangasi I, Braa K. The emergence of mobile-supported national health information systems in developing countries. MEDINFO 2010:540-544. [doi: 10.3233/978-1-60750-588-4-540]
19. Bishop AP, Mehra B, Bazzell I, Smith C. Participatory action research and digital libraries: Refraining evaluation. In: Digital Library Use: Social Practice in Design and Evaluation. Cambridge, Massachusetts: The MIT Press; 2003:161-190.
20. Hansen SK, Wilson LS, Robertson T. Applying an integrated approach to the design, implementation and evaluation of a telemedicine system. J Int Soc Telemed eHealth 2013;1(1):19-29 [FREE Full text]
21. Heffernan KJ, Chang S, Maclean ST, Callegari ET, Garland SM, Reavley NJ, et al. Guidelines and recommendations for developing interactive eHealth apps for complex messaging in health promotion. JMIR Mhealth Uhealth 2016 Feb;4(1):e14 [FREE Full text] [doi: 10.2196/mhealth.4423] [Medline: 26860623]
22. Katibeh M, Kalantarion M, Sabbaghi H, Mousavi B, Schriver M, Nikkhah H, et al. Designing a screening program for prevention of avoidable blindness in Iran through a participatory action approach. J Ophthalmic Vis Res 2019 Jan;14(1):52-61 [FREE Full text] [doi: 10.4103/jovr.jovr_43_18] [Medline: 30820288]
23. Lehto P. Robots with and for the elderly people: case study based on action research. In: ICERI2017 Proceedings. 2017 Presented at: 10th annual International Conference of Education, Research and Innovation; November 16-18, 2017; Seville p. 381-387. [doi: 10.21125/iceri.2017.0153]
24. Molapo M, Densmore M, Morie L. Designing with community health workers: enabling productive participation through exploration. In: Proceedings of the First African Conference on Human Computer Interaction. 2016 Nov Presented at: AfriCHI'16: African Conference for Human Computer Interaction; November 21-25, 2016; Nairobi p. 58-68. [doi: 10.1145/2998581.2998589]
25. Raij K. Caring TV as a forerunner in developing eHealth and eWelfare services. Gerontechnology 2016;15(2):130-145. [doi: 10.4017/gt.0000.00.00.000.00]
26. Rendón A, Martínez A, Dulcey MF, Seoane J, Shoemaker RG, Villarroel V, et al. Rural telemedicine infrastructure and services in the Department of Cauca, Colombia. Telemed J E Health 2005 Aug;11(4):451-459. [doi: 10.1089/tmj.2005.11.451] [Medline: 16149891]
27. Rönkkö K. An activity tracker and its accompanying app as a motivator for increased exercise and better sleeping habits for youths in need of social care: field study. JMIR Mhealth Uhealth 2018 Dec;6(12):e193 [FREE Full text] [doi: 10.2196/mhealth.9286] [Medline: 30578186]
28. Taylor J, Coates E, Wessels B, Mountain G, Hawley MS. Implementing solutions to improve and expand telehealth adoption: participatory action research in four community healthcare settings. BMC Health Serv Res 2015 Dec;15(1):529 [FREE Full text] [doi: 10.1186/s12913-015-1195-3] [Medline: 26626564]
29. Thobias J, Kiwanuka A. Design and implementation of an m-health data model for improving health information access for reproductive and child health services in low resource settings using a participatory action research approach. BMC Med Inform Decis Mak 2018 Jun;18(1):45 [FREE Full text] [doi: 10.1186/s12911-018-0622-x] [Medline: 29941008]
30. Callén B, Domènech AM, López GD, Tirado SFJ. Telecare research: (cosmo)politicizing methodology. Alter 2009 Apr;3(2):110-122. [doi: 10.1016/j.alter.2009.02.001]
31. VanHeerwaarden N, Ferguson G, Abi-Jaoude A, Johnson A, Hollenberg E, Chaim G, et al. The optimization of an eHealth solution (thought spot) with transition-aged youth in postsecondary settings: participatory design research. J Med Internet Res 2018 Mar;20(3):e79 [FREE Full text] [doi: 10.2196/jmir.8102] [Medline: 29510970]
32. Byrne E, Sahay S. Participatory design for social development: a South African case study on community-based health information systems. Inf Technol Dev 2010 Mar;13(1):71-94. [doi: 10.1002/itdj.20052]
33. Dansky KH, Bowles KH, Britt T. Nurses' responses to telemedicine in home healthcare. J Healthc Inf Manag 1999;13(4):27-38. [Medline: 10747699]
34. Chipps J, Ramlall S, Madigoe T, King H, Mars M. Developing telepsychiatry services in KwaZulu-Natal—an action research study. Afr J Psychiatry 2012 Jul;15(4):255-263. [doi: 10.4314/ajpsy.v15i4.33] [Medline: 22829228]
35. Day K, Kenealy TW, Sheridan NF. Should we embed randomized controlled trials within action research: arguing from a case study of telemonitoring. BMC Med Res Methodol 2016 Jun;16(1):1-11. [doi: 10.1186/s12874-016-0175-6] [Medline: 27277940]
36. Fennell KM, Turnbull DA, Bidargaddi N, McWha JL, Davies M, Olver I. The consumer-driven development and acceptability testing of a website designed to connect rural cancer patients and their families, carers and health professionals with appropriate information and psychosocial support. Eur J Cancer Care (Engl) 2017 Sep;26(5):e12533. [doi: 10.1111/ecc.12533] [Medline: 27405399]
37. Gaur A, Osella M, Ferro E, Hedman J. Open innovation as business model game-changer in the public sector. In: ECIS 2017 Proceedings. Atlanta: Association for Information Systems. AIS Electronic Library (AISeL); 2017:2317-2330.
38. Hansen SK, Robertson T, Wilson L, Hall R. Using an action research approach to design a telemedicine system for critical care: a reflection. Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat 2008 Dec:255-258. [doi: 10.1145/1517744.1517767]

39. Andersen T, Bjørn P, Kensing F, Moll J. Designing for collaborative interpretation in telemonitoring: re-introducing patients as diagnostic agents. Int J Med Inform 2011 Aug;80(8):e112-e126. [doi: 10.1016/j.ijmedinf.2010.09.010] [Medline: 21067968]
40. Webb MJ, Wadley G, Sanci LA. Experiences of general practitioners and practice support staff using a health and lifestyle screening app in primary health care: implementation case study. JMIR Mhealth Uhealth 2018 Apr;6(4):e105 [FREE Full text] [doi: 10.2196/mhealth.8778] [Medline: 29691209]
41. Alexander T, Singh N, Huda S, Alarakhia M. Project ALIVE: an action-research exploration of EMR value in primary care. Stud Health Technol Inform 2015;208:27-34. [doi: 10.3233/978-1-61499-488-6-27] [Medline: 25676942]
42. Trondsen M, Sandaunet AG. The dual role of the action researcher. Eval Program Plann 2009 Feb;32(1):13-20 [FREE Full text] [doi: 10.1016/j.evalprogplan.2008.09.005] [Medline: 19019435]
43. Park A, Chang H, Lee KJ. Action research on development and application of Internet of Things services in hospital. Healthc Inform Res 2017 Jan;23(1):25-34 [FREE Full text] [doi: 10.4258/hir.2017.23.1.25] [Medline: 28261528]
44. Street AF, Swift K, Annells M, Woodruff R, Gliddon T, Oakley A, et al. Developing a web-based information resource for palliative care: an action-research inspired approach. BMC Med Inform Decis Mak 2007 Sep;7(1):26 [FREE Full text] [doi: 10.1186/1472-6947-7-26] [Medline: 17854509]
45. Norman CD, Skinner HA. Engaging youth in e-health promotion: lessons learned from a decade of TeenNet research. Adolesc Med State Art Rev 2007 Aug;18(2):357-369. [Medline: 18605651]
46. Day K, Kerr P. The potential of telehealth for 'business as usual' in outpatient clinics. J Telemed Telecare 2012 Apr;18(3):138-141. [doi: 10.1258/jtt.2012.SFT104] [Medline: 22362833]
47. Simm W, Ferrario MA, Gradinar A, Tavares Smith M, Forshaw S, Smith I, et al. Anxiety and autism: towards personalized digital health. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems. 2016 Presented at: CHI'16: CHI Conference on Human Factors in Computing Systems; May 7-12, 2016; San Jose, CA, United States p. 1270-1281. [doi: 10.1145/2858036.2858259]
48. Phanareth K, Vingtoft S, Christensen AS, Nielsen JS, Svenstrup J, Berntsen GKR, et al. The epital care model: a new person-centered model of technology-enabled integrated care for people with long term conditions. JMIR Res Protoc 2017 Jan;6(1):e6 [FREE Full text] [doi: 10.2196/resprot.6506] [Medline: 28093379]
49. Lundberg N, Wintell M, Lindsköld L. The future progress of teleradiology-an empirical study in Sweden. Eur J Radiol 2010 Jan;73(1):10-19. [doi: 10.1016/j.ejrad.2009.10.015] [Medline: 20022724]
50. Stensæth K, Ruud E. An interactive technology for health: new possibilities for the field of music and health and for music therapy? A case study of two children with disabilities playing with 'ORFI'. Music, Health, Technology and Design 2014;8:39-66 [FREE Full text]
51. Waterman H, Marshall M, Noble J, Davies H, Walshe K, Sheaff R, et al. The role of action research in the investigation and diffusion of innovations in health care: the PRIDE project. Qual Health Res 2007 Mar;17(3):373-381. [doi: 10.1177/1049732306298976] [Medline: 17301345]
52. Berger G, Peerson A. Giving young Emirati women a voice: participatory action research on physical activity. Health Place 2009 Mar;15(1):117-124. [doi: 10.1016/j.healthplace.2008.03.003] [Medline: 18515171]
53. Holeman I, Kane D. Human-centered design for global health equity. Inf Technol Dev 2019 Sep;26(3):477-505 [FREE Full text] [doi: 10.1080/02681102.2019.1667289] [Medline: 32982007]
54. Kimaro H, Twaakyondo H. Analysing the hindrance to the use of information and technology for improving efficiency of health care delivery system in Tanzania. Tanzan Health Res Bull 2005 Sep;7(3):189-197. [doi: 10.4314/thrb.v7i3.14259] [Medline: 16941947]
55. Baskerville R, Myers MD. Special issue on action research in information systems: making IS research relevant to practice: foreword. MIS Quarterly 2004 Sep;28(3):329-335. [doi: 10.2307/25148642]
56. Baum F, MacDougall C, Smith D. Participatory action research. J Epidemiol Community Health 2006 Oct;60(10):854-857 [FREE Full text] [doi: 10.1136/jech.2004.028662] [Medline: 16973531]
57. Trist L. Action research and adaptive planning. In: Experimenting With Organizational Life. Boston: Springer; 1976:223-236.
58. Freire P. Pedagogy of the Oppressed. New York: Continuum International Publishing Group; 1970.
59. Reason P, McArdle K. Brief notes on the theory and practice of action research. In: Becker S, Bryman A, editors. Understanding Research Methods for Social Policy and Practice. London: The Polity Press; 2004:114-122.
60. Bradbury-Huang H. What is good action research? Action Res 2010 Mar;8(1):93-109. [doi: 10.1177/1476750310362435]
61. Rapoport RN. Three dilemmas in action research. Human Relations 2016 Apr;23(6):499-513. [doi: 10.1177/001872677002300601]
62. Avison DE, Lau F, Myers MD, Nielsen PA. Action research. Commun ACM 1999 Jan;42(1):94-97. [doi: 10.1145/291469.291479]
63. Baskerville RL, Wood-Harper AT. A critical perspective on action research as a method for information systems research. J Inf Technol 1996 Sep;11(3):235-246. [doi: 10.1080/026839696345289]
64. Baskerville RL. Investigating information systems with action research. CAIS 1999 Oct;2:19. [doi: 10.17705/1CAIS.00219]
65. Braa J, Monteiro E, Sahay S. Networks of action: sustainable health information systems across developing countries. MIS Quarterly 2004 Sep;28(3):337-362. [doi: 10.2307/25148643]

66. Kemmis S, McTaggart R, Nixon R. The Action Research Planner. Singapore: Springer; 2014.
67. Denzin NK, Lincoln YS, editors. Strategies of Qualitative Inquiry. Thousand Oaks: Sage; 2008.
68. Dick B. Review essay: Utopia made practical? Action research comes of age. Forum: Qualitative Social Research 2002;3(1):1-2. [doi: 10.17169/fqs-3.1.892]
69. Sankaran S, Dick B, Passfield R, Swepson P, editors. Effective Change Management Using Action Learning and Action Research: Concepts, Frameworks, Processes, Applications. Cornell: Southern Cross University Press; 2001:225-229.
70. Elden M, Levin M. Cogenerative learning: bringing participation into action research. In: Participatory Action Research. London: Sage; 1991:127-142.
71. Robson C. Real world research: A Resource for Social Scientists and Practitioner–Researchers. Oxford: Blackwell; 2002.
72. Skinner HA, Maley O, Norman CD. Developing internet-based eHealth promotion programs: the Spiral Technology Action Research (STAR) model. Health Promot Pract 2006 Oct;7(4):406-417. [doi: 10.1177/1524839905278889] [Medline: 16840770]
70. Skinner HA. Promoting Health Through Organizational Change. San Francisco: Pearson; 2002.
74. Susman GI, Evered RD. An assessment of the scientific merits of action research. Administrative Science Quarterly 1978 Dec;23(4):582-603. [doi: 10.2307/2392581]
75. Hart E, Bond M. Developing action research in nursing. Nurse Researcher 1995 Mar;2(3):4-14. [doi: 10.7748/nr.2.3.4.s2]
76. Hart E, Anthrop C. Action research as a professionalizing strategy: issues and dilemmas. J Adv Nurs 1996 Mar;23(3):454-461. [doi: 10.1111/j.1365-2648.1996.tb00006.x] [Medline: 8655819]
77. Ferrario MA, Simm W, Forshaw S, Gradinar A, Smith MT, Smith I. Values-first SE: research principles in practice. In: IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C). 2016 May Presented at: IEEE/ACM 38th International Conference on Software Engineering Companion (ICSE-C); May 14-22, 2016; Austin p. 553-562. [doi: 10.1145/2889160.2889219]
78. Baldwin JN, Napier S, Neville S, Wright-St Clair VA. Impacts of older people's patient and public involvement in health and social care research: a systematic review. Age Ageing 2018 Nov;47(6):801-809. [doi: 10.1093/ageing/afy092] [Medline: 29939208]
79. Harrison JD, Auerbach AD, Anderson W, Fagan M, Carnie M, Hanson C, et al. Patient stakeholder engagement in research: a narrative review to describe foundational principles and best practice activities. Health Expect 2019 Jun;22(3):307-316 [FREE Full text] [doi: 10.1111/hex.12873] [Medline: 30761699]
80. Tomlinson SRL, Gore N, McGill P. Training individuals to implement applied behavior analytic procedures via telehealth: a systematic review of the literature. J Behav Educ 2018 Mar;27(2):172-222. [doi: 10.1007/s10864-018-9292-0]
81. McNeil H, Elliott J, Huson K, Ashbourne J, Heckman G, Walker J, et al. Engaging older adults in healthcare research and planning: a realist synthesis. Res Involv Engagem 2016 Mar;2(1):10 [FREE Full text] [doi: 10.1186/s40900-016-0022-2] [Medline: 29062511]
82. Brydon-Miller M, Aragón AO. The 500 hats of the action researcher. In: Bilfeldt A, Jørgensen MS, Andersen J, Perry KA, editors. Den ufærdige fremtid – Aktionsforskningens potentialer og udfordringer. Denmark: Aalborg Universitetsforlag; 2018:19.
83. Canto-Farachala P, Larrea M. Rethinking the communication of action research: Can we make it dialogic? Action Research 2020 Feb:147675032090589. [doi: 10.1177/1476750320905896]

Abbreviations

ADR: action design research
AR: action research
HCI: human computer interaction
PAR: participatory action research
PD: participatory design

Edited by A Mavragani; submitted 05.07.21; peer-reviewed by S Lifvergren, M Bestek, R Nuijten; comments to author 13.08.21; revised version received 09.09.21; accepted 06.12.21; published 28.01.22.

Please cite as:
Oberschmidt K, Grünloh C, Nijboer F, van Velsen L
Best Practices and Lessons Learned for Action Research in eHealth Design and Implementation: Literature Review
J Med Internet Res 2022;24(1):e31795
URL: https://www.jmir.org/2022/1/e31795
doi:10.2196/31795
PMID:35089158

Review

Digital Health Promotion and Prevention in Settings: Scoping Review

Anna Lea Stark[1*], MSc; Cornelia Geukes[1*], DrPH; Christoph Dockweiler[2], Prof Dr

[1]School of Public Health, Centre for ePublic Health, Bielefeld University, Bielefeld, Germany

[2]Faculty V: School of Life Sciences, University of Siegen, Siegen, Germany

[*]these authors contributed equally

Corresponding Author:
Anna Lea Stark, MSc
School of Public Health
Centre for ePublic Health
Bielefeld University
Universitätsstraße 25
Bielefeld, 33615
Germany
Phone: 49 521 106 67859
Email: lea.stark@uni-bielefeld.de

Abstract

Background: Digital technologies are increasingly integrating into people's daily living environments such as schools, sport clubs, and health care facilities. These settings play a crucial role for health promotion and prevention because they affect the health of their members, as the World Health Organization has declared. Implementing digital health promotion and prevention in settings offers the opportunity to reach specific target groups, lower the costs of implementation, and improve the health of the population. Currently, there is a lack of scientific evidence that reviews the research on digital health promotion and prevention in settings.

Objective: This scoping review aims to provide an overview of research targeting digital health promotion and primary prevention in settings. It assesses the range of scientific literature regarding outcomes such as applied technology, targeted setting, and area of health promotion or prevention, as well as identifies research gaps.

Methods: The scoping review was conducted following the Levac, Colquhoun, and O'Brien framework. We searched scientific databases and gray literature for articles on digital setting–based health promotion and prevention published from 2010 to January 2020. We included empirical and nonempirical publications in English or German and excluded secondary or tertiary prevention and health promotion at the workplace.

Results: From 8888 records, the search resulted in 200 (2.25%) included publications. We identified a huge diversity of literature regarding digital setting–based health promotion and prevention. The variety of technology types extends from computer- and web-based programs to mobile devices (eg, smartphone apps) and telemonitoring devices (sensors). We found analog, digital, and blended settings in which digital health promotion and prevention takes place. The most frequent analog settings were schools (39/200, 19.5%) and neighborhoods or communities (24/200, 12%). Social media apps were also included because in some studies they were defined as a (digital) setting. They accounted for 31.5% (63/200) of the identified settings. The most commonly focused areas of health promotion and prevention were physical activity (81/200, 40.5%), nutrition (45/200, 22.5%), and sexual health (34/200, 17%). Most of the interventions combined several health promotion or prevention methods, including environmental change; providing information, social support, training, or incentives; and monitoring. Finally, we found that the articles mostly reported on behavioral rather than structural health promotion and prevention.

Conclusions: The research field of digital health promotion and prevention in settings is heterogeneous. At the same time, we identified research gaps regarding the absence of valid definitions of relevant terms (eg, digital settings) and the lack of literature on structural health promotion and prevention in settings. Therefore, it remains unclear how digital technologies can contribute to structural (or organizational) changes in settings. More research is needed to successfully implement digital technologies to achieve health promotion and prevention in settings.

(J Med Internet Res 2022;24(1):e21063) doi:10.2196/21063

KEYWORDS

setting approach; health promotion; health prevention; eHealth; internet; behavior change; web-based intervention; technology; mobile phone

Introduction

Background

The setting approach is considered a main strategy of health promotion [1]. For its practical implementation, it is important to use an operationalizable and consensual concept of a setting. However, various and sometimes divergent definitions exist [2]. A setting can be defined as a relatively permanent social context of which the members are subjectively aware [3] or as a "...place or social context in which people engage in daily activities in which environmental, organizational and personal factors interact to affect health and wellbeing" [4]. Up to now, there has been a lack of a conceptual and theoretical framework of the setting approach [2]. Health-promoting organizational development therefore often makes use of sociological organizational theories, especially systems theory and structuration theory. Furthermore, organizational change theories are applied, including, for example, the diffusion theory of Everett Rogers or the model of change proposed by Kurt Lewin [5]. Public health theories and models are also used for the implementation of setting-based or complex health-related interventions partly considering structural aspects of the implementation, such as the PRECEDE–PROCEED model [6].

To gain a better understanding of what setting-based health promotion and prevention might look like in practice, Whitelaw et al [7] developed a categorization of 5 different types. The passive model focuses on individuals and their health-relevant behavior. The setting is used as an access to individuals and groups. The active model includes interventions that focus on the health behavior of individuals. However, it assumes that some behavior changes are inhibited or enabled by structural features of the setting. Thus, structural modifications are required. The vehicle model focuses on individual health issues, but interventions and projects are intended to initiate systemic changes. The projects serve as *door openers* for developments at other levels of the setting. The focus is on structural factors and their influence on health, even if individual health issues are addressed first. In the organic model, social systems are understood as the product of the processes of individuals and groups living and acting together. Interventions result in systemic changes that affect the communication culture and community values in the setting. The setting is understood as the framework for health, which can be influenced by setting members. Participation and empowerment of people to affect their environment are key concepts. The comprehensive model assumes that the setting affects health. Settings are understood as superordinate systems in which the individual has hardly any opportunities to affect them himself. Interventions are often implemented from the top down [1,7].

In the last decade, digitization has had a major impact on everyday life in settings and on the research and practice of setting-based health promotion and prevention. Approximately 77% of the population uses the internet on a daily basis, and approximately 90% of the population uses it occasionally [8]. Digital media are the most important source of information, especially among the younger population. With the development of health-related digitization (also known as eHealth), internet-based health information and the use of mobile apps for health topics (also known as mobile health [mHealth]) have made extreme progress and replaced traditional face-to-face sources of information [9-11]. Digitization offers the opportunity to customize health promotion and prevention interventions and adapt them to the needs of individuals, for example, through personalized motivational SMS text messages or reminders. Besides 1-way communication of health information through mass media (eg, videos or e-books), digital health promotion and prevention can be realized with interaction in the form of web-based games or web-based forums. Other eHealth possibilities involve the monitoring of health data, web-based training or classes, digital consultations with health experts, or the provision of health information through (video) games, referred to as serious games [12]. At the same time, settings, for example, organizations and institutions, are facing digital transformation processes concerning structure, management, and organizational development, and new (digital) types of organizations are emerging, for example, virtual organizations [13]. In the health sector, digital transformation affects various organizations such as hospitals (*Hospital 4.0*) and pharmacies (*web-based pharmacies*) [14]. With the increased use of eHealth and mHealth apps, everyday life in general is becoming more and more digitalized, which has various effects on health. An example is the digital gaming environment, where web-based players face new health risks as well as accrue new health benefits [15]. Overall, the digitization and the large number of available digital technologies offer great potential in terms of health promotion and prevention, especially in settings.

However, there is still a lack of a clear and consensual definition for digital technologies in health promotion and prevention on the one hand [16] and a lack of a systematic description of the extent and characteristics of this research field on the other hand. Only a few reviews show the range of this topic and illustrate the urgency of scientific work on this issue [17,18].

As we have stated previously, settings are an important starting point for health promotion and prevention efforts because of their huge impact on people's health [19] and they can be used for digital health promotion and prevention, but this new research field has not been systematically analyzed yet [20-23]. To effectively use the full potential of digital technologies, it is important to define the framework in which settings can be achieved, changed, and used in the field of digital health promotion and prevention [24]. Therefore, it might be helpful to analyze digital interventions in settings according to existing theoretical frameworks that are known from analog setting-based interventions, for example, the types of setting-based health promotion outlined by Whitelaw et al [7]. There is also an urgent need to identify which (digital) structures and factors in organizations need to be changed to achieve successful digital

health promotion [5]. By defining and structuring the field of digital health promotion and prevention in settings, the quality and quantity of digital interventions can improve and increase, respectively, in the future.

Objectives

The overall objective of this scoping review is to provide an overview of the research targeting digital health promotion and primary prevention in settings. More specifically, it aims to (1) assess the range of scientific literature focusing on digital health promotion and primary prevention in settings regarding various outcomes such as applied technology, targeted setting or living environment (these terms are used synonymously), and area of health promotion and prevention and (2) identify research gaps resulting from these findings.

Methods

This paper follows the reporting guidelines of the PRISMA-ScR (Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews) checklist [25].

Scoping Review Methodology

Overview

We conducted a scoping review that aims at mapping a broad body of literature, identifying research gaps, and making recommendations for future research [26,27]. We followed the 6-step framework outlined by Levac et al [28], who updated and extended the initial framework developed by Arksey and O'Malley [27]. The steps are described separately in the following sections.

Step 1: Identifying the Research Questions

The research questions were developed by 2 reviewers (ALS and CG) through an iterative process that included exploratory literature searches, an exploration of relevant terms and concepts, and thematic discussions. In addition, we interviewed experts in the field of scientific reviews in public health on the thematic content, methodological approach, and research questions.

Step 2: Identifying Relevant Studies

The search strategy was developed with the support of a specialized librarian in public health. The following scientific databases were searched:

- MEDLINE (through PubMed)
- CINAHL (through EBSCO)
- SocINDEX (through EBSCO)
- PsycINFO (through EBSCO)
- PSYNDEX (through EBSCO)
- IEEE Xplore
- BASE
- Web of Science

Search terms were identified for the categories *health promotion and prevention*, *living environments and settings*, and *technology and digital*, including Medical Subject Headings terms and other index terms, as well as synonyms. Search terms were combined using the Boolean operators *OR* and *AND*. The search was restricted to articles published in English or German from 2010 to January 2020. To exclude reviews and web-based surveys (that do not report on a technology for health promotion or prevention) the terms *NOT review* and *NOT online survey* were used. Where possible, the database filter *humans* was applied. The exclusion criteria are described in detail in the next section. We also included gray literature. Therefore, we screened websites focusing on the topic of digitization and health promotion or prevention. Websites of governments, health promotion associations, medical or health-related facilities, health insurance companies, and universities and other scientific institutions were screened. The last research step included a screening of the reference lists of all eligible articles. The full search strategy, including the scientific databases and search terms, is provided in Multimedia Appendix 1.

Step 3: Study Selection

Overview

The references were independently screened by 2 reviewers (ALS and CG) for consistency with predefined inclusion and exclusion criteria, starting with screening the titles and abstracts and followed by screening the full texts. Rayyan (Qatar Computing Research Institute), a web-based application developed to facilitate the review process, was used for this step. Conflicts were discussed between the reviewers (ALS and CG) and with scientific experts according to the inclusion and exclusion criteria. The a priori–defined inclusion and exclusion criteria are described in the next sections using the Population, Concept, and Context principle [29].

Concept

The included articles had to report on interventions for health promotion or primary prevention. Primary prevention includes all those interventions that are implemented before the first occurrence of an undesirable condition. Examples are vaccinations for people who do not belong to a risk group or interventions to prevent unhealthy behavior or organizational health risks. Primary prevention must be differentiated from secondary prevention aiming at the early detection or control of diseases and from tertiary prevention targeting the chronification, consequential damage, or relapse of diseases. Methods for disease prevention could be education, normative and regulative proceedings, or economic incentive and sanction mechanisms [30]. All articles concerning secondary or tertiary prevention were excluded. For this reason, we excluded articles focusing on already present risk behaviors such as smoking cessation or weight loss for people who are overweight, as well as articles addressing screenings for early disease detection. Furthermore, a distinction can be made among universal, selective, and indicated prevention. We included universal prevention; for example, if an intervention was implemented in an entire school class with the aim of health promotion or primary prevention. We excluded selective and indicated interventions, which are defined as interventions focusing on people with risk factors or precursors of a disease [30]. We included preventive or health-promoting interventions at specialists' practices (eg, sexual health clinics) if the intervention was not aimed in particular at people who are already ill or those classified as at risk. Thus, it is possible that healthy people

attend a preventive health screening or accompany someone who is ill. Furthermore, it is possible that the health issue that led to their appointment is separate from the primary preventive or health-promoting intervention. Besides primary prevention, articles were included if they reported on health promotion. Health promotion interventions aim at enabling the population to enhance their health by strengthening resources [19]. Interventions focus either on strengthening individuals to be empowered or on augmenting the social, ecological, and economic factors that influence people's health [31]. In this context, the concept of *healthy settings* has to be mentioned. According to the World Health Organization (WHO), a setting for health is a "place or social context in which people engage in daily activities in which environmental, organizational and personal factors interact to affect health and wellbeing" [4,32]. Rosenbrock [3] follows a broader approach and defines settings as a relatively permanent social context that the members have to be subjectively aware of. This includes rather informal settings such as neighborhoods or groups of people with the same value orientation or life circumstances. In our review, we want to outline the research field in its whole breadth; hence, we will refer to the broader setting definition. This means that we also include more informal settings.

In addition, the included articles had to involve up-to-date digital technologies that are used by the setting members. Digital technologies differ from analog ones in that they "operate on the basis of a discrete code. Usually this code is binary, constructed from sequences of binary digits or bits, each of which can be in one of two states, named 0 and 1" [33]. Therefore, we included web-based programs and mobile apps that are used through a computer, telephone, smartphone, tablet, or other digital devices. Games played on a gaming console were included too. Interventions using videos only were included if these were available through a mobile app (eg, tablet) or were accessible on the web. Interventions using traditional or, rather, obsolete DVDs, television, or radio were excluded.

Population

Following the definition of health promotion and primary prevention, articles were included if the intervention targeted healthy people. Therefore, articles focusing on people engaging in risky health behavior or already having symptoms or diseases were excluded.

Context

Health promotion or prevention interventions were included if they referred to a setting according to the definitions suggested by the WHO [4] and Rosenbrock [3]. Examples are formal organizations such as schools, universities, and health care facilities. Furthermore, we included less formal settings such as households and neighborhoods because they meet the criteria of a setting by featuring a social context that the setting members (eg, household members or neighbors) are aware of. Articles relating to the workplace as a setting were excluded because the field of health promotion and prevention at work is a separate specific research area within public health. The reviewers had the presupposition that the definitions and characteristics of formal and informal settings might be applicable to digital environments or a digital social context (eg, social media platforms). Therefore, the accordance of the screened articles with the definition of a setting was checked not only for analog conditions, but also for digital conditions. According to the definitions, a criterion is the existence of an accepted social system, namely a given social context that the users of the digital technology are aware of. This criterion was considered fulfilled if the digital technology allowed social interaction among different users of the technology and if the users were aware of the existence of other users. Another criterion was that the social context exists relatively permanently. This criterion was considered fulfilled if the social interaction within the digital technology was not only possible once, but also over a longer period. In accordance with the definition, the identified digital settings could be characterized as formal (such as digital organizations) or informal (such as digital groups of friends).

Besides the Population, Concept, and Context principle, another inclusion criterion concerned the publication type and study design. Nearly all publication types, empirical as well as nonempirical (eg, discussion papers), and study designs were included. The only study design excluded was reviews because the studies included in reviews would have been listed in the relevant databases and would have been identified through our search strategy if they corresponded to our inclusion criteria. In addition, we excluded abstracts and posters because they would not deliver all the required study information. Furthermore, articles had to be published in English or German from 2010 to January 2020. The time restriction was necessary because of the rapid and significant changes that occur in the digital sector.

Step 4: Charting the Data

Data extraction was conducted independently by 2 reviewers (ALS and CG). For this purpose, a data extraction form was developed, which was discussed during the expert consultations and updated during the extraction process. The reviewers extracted general publication characteristics, for example, year of publication, country, and publication type or study design on the one hand, as well as specific outcomes that are relevant for the description and mapping of the research field, for example, the type of applied technology, the type of setting, the target group of the intervention, and the area and method of health promotion or prevention on the other hand.

The distinction according to the technology type was made based on the study by O'Neil et al [34], which included mobile devices (eg, smartphones and tablets), computer- and web-based programs, social media apps, and telemonitoring devices. We added gaming consoles to the classification. The method of health promotion and prevention was determined according to the distinction by Leppin [30]. The author first distinguishes whether the behavior or the environment is being changed. Second, the author differentiates among educational interventions (including information provision and training), normative and regulative proceedings (eg, new laws or prohibitions), and economic incentive and sanction mechanisms. To this categorization, Scherenberg [12] adds methods for digital health prevention, including monitoring, reminder, social interaction, and social support (eg, digital challenges or encouragement).

Step 5: Collating, Summarizing, and Reporting the Results

Following Levac et al [28], this stage was divided into 3 distinct steps. First, we analyzed the results, including numerical summary analysis. Second, we reported the results focusing on the outcomes that referred to the overall research aims. Third, we considered and discussed the findings and their meaning for future research, practice, and policy. We did not carry out a critical appraisal of the methodological quality of the included articles because the aim of this scoping review is to provide an overview of the research field and not to analyze the effectiveness of interventions.

Step 6: Consultation

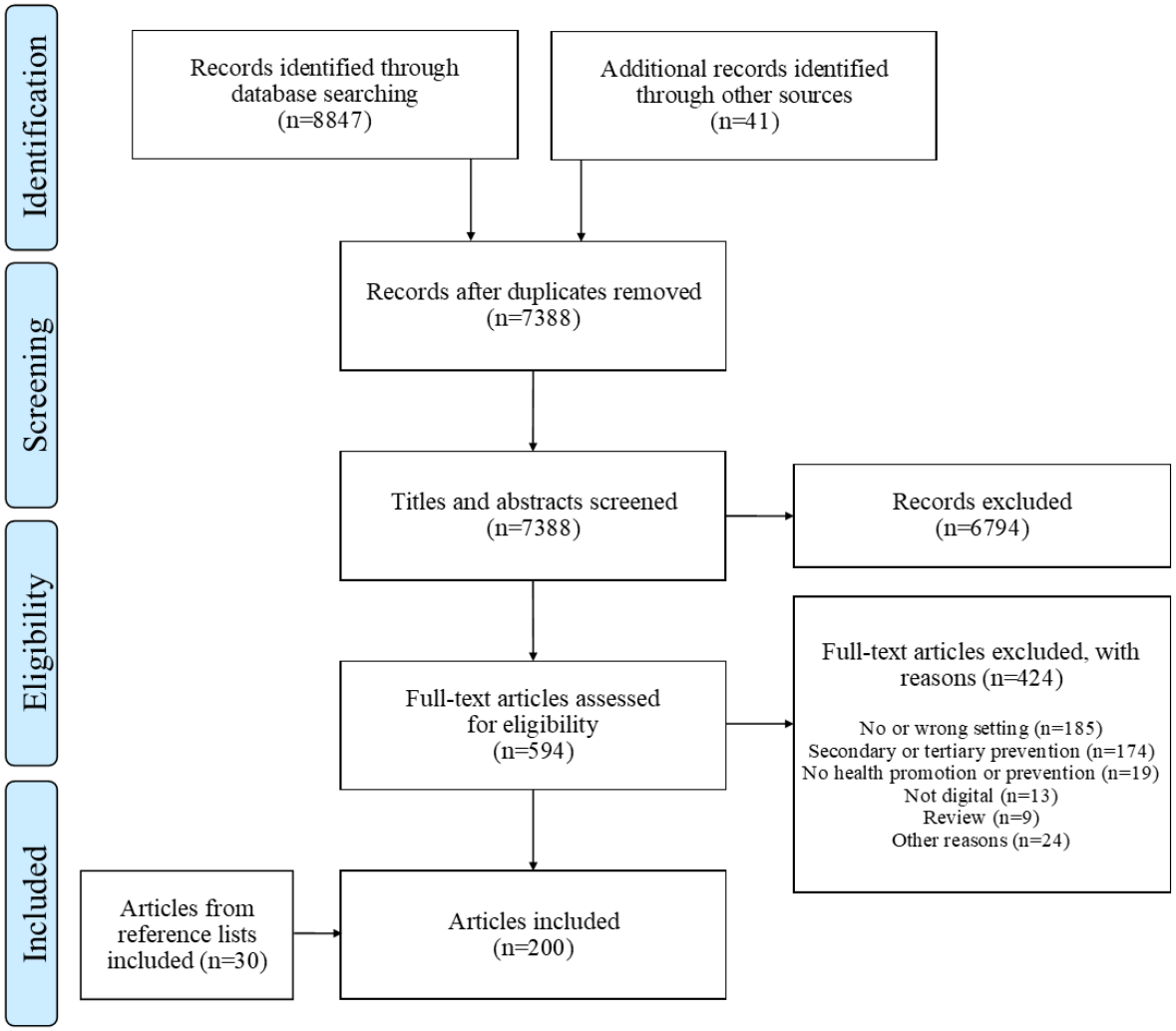
As mentioned previously, the review process was supported by scientific experts through regular consultations and discussions. The scientific advisory board included a specialized librarian, an expert in conducting literature reviews, and an expert in the field of digital health promotion and prevention.

Results

Overview

The initial database search resulted in the identification of 8847 records, and the gray literature search resulted in the identification of 41 records. From the 8888 articles, 1500 (16.88%) duplicates were removed, leaving 7388 (83.12%) articles eligible for screening by title and abstract. Of these 7388 articles, 6794 (91.96%) were excluded after the screening, leaving 594 (8.04%) articles for the full-text appraisal. Of these 594 articles, after the full-text appraisal, a further 424 (71.4%) were excluded, leaving 170 (28.6%) articles that met the inclusion criteria. In addition, the reference lists of these articles were screened, which led to a further 30 articles being identified; therefore, the literature search yielded 200 articles. The whole review process is summarized in the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) flowchart [35] in Figure 1.

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



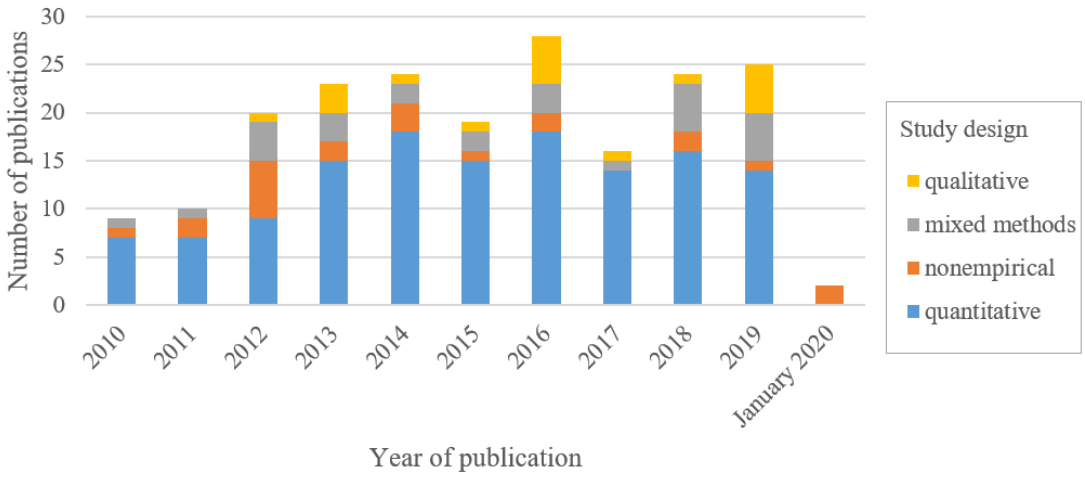
General Characteristics of the Publications

The articles included in this review were published from 2010 to January 2020. There were 9-28 publications per year in the research area of digital health promotion and prevention in settings (except for 2020 where the search was restricted to January of that year). Whereas the number of publications was 9 and 10 in 2010 and 2011, respectively, it doubled in 2012 and thereafter fluctuated from 16 to 28 publications, with the peak in 2016 and the minimum in 2017. The data extraction chart, including all extracted data used in this review, is provided in Multimedia Appendix 2 [36-235].

In total, we found publications from 25 different countries on 6 continents (based on the country where the first author's research institution is based). Of the 200 articles, 94 (47%) were published in the United States and 3 (1.5%) in Canada, whereas 96 (48%) were published in Europe. Of these 96 articles, 15 (16%) originated from Germany, 12 (13%) from the Netherlands, 11 (11%) from the United Kingdom, and 6 (6%) from Belgium, whereas other European countries, including Italy, Norway, Denmark, Portugal, Austria, Finland, Ireland, Sweden, Switzerland, Spain, and Greece, each accounted for 5 (5%) or fewer publications. Of the 200 articles, 21 (10.5%) were from Australia, whereas Asian countries, including India, Thailand, Taiwan, Korea, and China, each accounted for 3 (1.5%) or fewer publications; 2 (1%) articles originated from South America (Brazil), and 1 (0.5%) was published in an African country (Uganda).

Of the 200 articles, 178 (89%) were empirical studies. Most of these 178 studies had a quantitative study design: 65 (36.5%) were randomized controlled trials (RCTs), 22 (12.4%) were cross-sectional studies, 15 (8.4%) were pretest–posttest studies, 13 (7.3%) were secondary data analyses, 11 (6.2%) were study protocols, and 7 (3.9%) were controlled trials. Of the 178 empirical studies, a further 27 (15.2%) followed a mixed methods approach. Of these 27 studies, 4 (15%) were study protocols and 1 (4%) was a secondary data analysis. Of the 178 studies, 18 (10%) had a qualitative design. In addition, among the 200 studies, we found 22 (11%) nonempirical publications, including theoretical or discussion papers, technology or framework descriptions, and thematic overviews. Figure 2 shows the distribution of study designs by publication year.

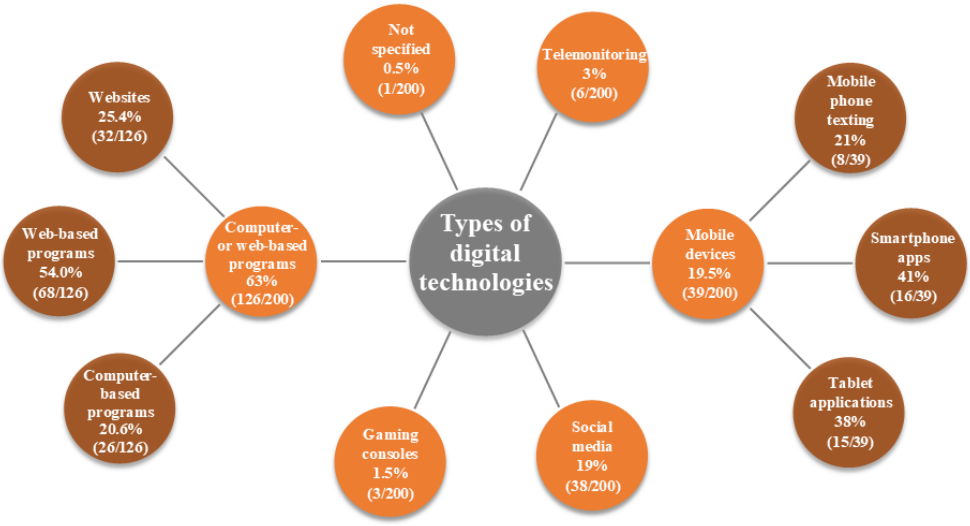
Figure 2. Study designs by publication year (n=200).



Type of Digital Technology Used for Heath Promotion and Prevention in Settings

As Figure 3 shows, the type of digital technology used for health promotion or prevention can be distributed through mobile devices (eg, smartphones and tablets), computer- and web-based programs, social media apps, telemonitoring devices (eg, external sensors or smartwatches), and gaming consoles. Of the 200 studies, 126 (63%) targeted computer- and web-based programs, 39 (19.5%) a mobile device, and 38 (19%) a social media app, whereas only a small number (6/200, 3%, or fewer) reported on one of the other technology types. As some publications reported on ≥1 technology types, the sum of the percentages adds up to >100% in >200 articles (this also applies to the targeted areas and methods of health promotion and prevention, described in the *Areas and Methods of Digital Health Promotion and Prevention in Settings* section).

Figure 3. Types of digital technologies (n=213 due to the fact that some publications report on more than one technology type).



Furthermore, these technology types offer different apps that are used for health promotion or prevention purposes. Regarding the technology type *mobile devices*, 41% (16/39) of the publications reported on smartphone apps and 38% (15/39) reported on tablet apps, whereas 21% (8/39) used mobile phone texting. Of the interventions that included a computer- or web-based program, 54.0% (68/126) were web-based programs, 25.4% (32/126) were websites, and 20.6% (26/126) were computer-based programs.

There were different apps and programs with a variety of digital functions and options for users mentioned in the publications, which cannot be described in detail in this review. A few apps are presented as examples here. An example is the use of a smartphone with an integrated geographic information system for the promotion of physical activity. The underlying idea is that specific environmental conditions of a neighborhood can promote physical activity. These features, referred to as neighborhood walkability, can be measured through a geographic information system and used to improve the environmental conditions [36]. Other interventions used computer-based programs, including virtual reality, for health promotion and prevention purposes. As an example, a study combined stationary cycling at independent living facilities for older adults with virtual reality to create a computer-simulated environment (cybercycling). This intervention, also referred to as an exergame (a digital game that requires one to perform physical exercise), aims at improving the cognitive function of older adults [37]. Another digital intervention concerns serious games, which are digital games with educational objectives played on computers or mobile devices. We identified serious games for various health promotion and prevention purposes, for example, a game tackling childhood obesity in schools [38] or a game on neglected tropical diseases developed for use in a community [39]. As an example for telemonitoring devices, a study focused on the real-time monitoring of children's physical activity in school measured by smartwatches and supervised by classroom teachers, which, over the course of time, enables teacher-regulated strategies for providing more opportunities for physical activity [40]. A gaming console was used, for example, by children in schools, where active video games were played to enhance their physical activity during school breaks [41]. Another intervention at a preschool investigated the use of a social media platform for the delivery of health-related education materials to augment a classroom-based obesity-prevention curriculum [42].

Target Groups and Targeted Settings in Digital Health Promotion and Prevention

The analysis identified 2 categories of target groups for digital health promotion and prevention interventions. First, there were interventions where the users of the technology were the direct target group for health promotion or prevention. Examples of direct target groups were adolescents, adults, children, church members, community members, expatriates or travelers, men, mothers or pregnant women, older adults, patients, school or university students, vacationers, women, and young adults. In addition, some publications focused on people from a specific cultural, linguistic, or sexual background (eg, refugees, bisexual men, and people living in rural areas). Others had >1 target group or did not specify the target group. Of the 200 included studies, most (48/200, 24%) targeted school students, followed by those targeting older adults (18/200, 9%), university students (17/200, 8.5%), and adults in general (15/200, 7.5%). Second, there were interventions where the technology users were not the direct target group for health promotion or prevention. Instead, the technology users (as multiplicators or mediators) were empowered or supported through the use of the technology to ensure that other people received health promotion or prevention. Examples of multiplicators or mediators were parents, teachers, doctors, and childcare center directors.

The aforementioned target groups already indicate some settings where digital health promotion and prevention interventions take place. The settings were divided into formal and informal settings. The formal analog settings that we identified were schools, universities, and other educational settings; community facilities (eg, churches or prisons); health care facilities (eg, primary care, dental, sexual health, or vaccine clinics); and care for older adults or childcare facilities. Informal analog settings were communities or neighborhoods and the household or home. We identified several digital technologies that might be

characterized as a digital setting at the time of publication. These were web-based distance learning universities; virtual reality apps; and social media, including apps, websites, and web-based platforms or web-based programs that offered discussion boards, forums, or other ways of social interaction. The web-based distance learning setting can be characterized as a formal digital setting.

Overall, with 31.5% (63/200) of the publications reporting on it, social media was the most frequent setting in digital health promotion and prevention. Another 19.5% (39/200) of the publications reported on schools as a setting, 12% (24/200) on neighborhoods or communities, and 6.5% (13/200) on health care facilities. Furthermore, 6% (12/200) focused on the informal setting *household or home* and 3% (6/200) on the formal setting *universities*. Other publications thematized interventions that used combinations of analog and digital settings that we refer to as blended settings. For instance, interventions in the community were often complemented with a digital app that itself represented a digital setting (eg, a social media platform). The same combinations were found for schools and universities.

Areas and Methods of Digital Health Promotion and Prevention in Settings

The included articles dealt with different areas of health promotion and prevention. Physical activity interventions were the most commonly applied, with 40.5% (81/200) of the publications thematizing them. Other often considered areas were nutrition in 22.5% (45/200) of the publications; sexual health in 17% (34/200); substance use, for example, alcohol or drug use, in 10% (20/200); weight in 9.5% (19/200); and mental health in 5.5% (11/200). Only a few publications described interventions for health promotion or well-being in general (9/200, 4.5%), for vaccination (8/200, 4%), and for various risk factors at once (10/200, 5%). Other health topics that were tackled rarely were bone and oral health, sound and sun exposure, cognitive function, fall prevention, sleep, and transmission of diseases (17/200, 8.5%; Multimedia Appendix 3).

Besides the areas of health promotion and prevention, we analyzed the methods used to achieve health promotion and prevention among the target group. Most of the publications reported on interventions that combined several of the following methods: environmental change; providing information, social support, training, or incentives; and monitoring. The provision of information was the most used method (181/200, 90.5%). It aimed at improving the knowledge of the target group regarding a specific health promotion or prevention topic and thereby changing health behavior. Of the 200 interventions, 79 (39.5%) offered some form of social support, for example, through connecting people with the same health problem or the same health goals to motivate the target group. Approximately every fifth intervention (41/200, 20.5%) used monitoring as a method for health promotion and prevention. By giving users the opportunity to collect, visualize, and control health-related data over a longer period, monitoring enabled the user to recognize progress and stay on track. Of the 200 articles, 30 (15%) aimed at changing environmental conditions, for example, a change in the supermarket environment [43] or the school environment [44]. Only a few interventions used incentives as a mechanism to influence health behavior (8/200, 4%) or provide physical or mental training (19/200, 9.5%).

The different methods of health promotion and prevention indicate that one should further distinguish between structural and behavioral health promotion and prevention in settings. Structural health promotion and prevention is realized by changing setting-based or organizational structures. Only a small percentage (30/200, 15%) of the identified articles dealt with structural health prevention. Of these 30 articles, 24 (80%) focused on environmental or organizational changes in analog settings such as neighborhoods, supermarkets, schools, childcare centers, and home environments. An example is a smartphone app that enabled community members to document neighborhood features through geocoded photographs. The intervention aimed at identifying and changing aspects of the physical and social environment that influenced the health behavior of the community members [45]. Only a few articles (6/30, 20%) discussed structural changes in digital settings, for example, social networking sites.

Behavioral health promotion and prevention interventions were used in 85% (170/200) of the included articles. In these behavioral interventions, the setting served as a place to reach a specific target group and offer them an intervention for changing their health behavior. The provision of a tablet in the waiting room of a sexual health clinic that provided access to a website with information for safer sex aiming at a change in sexual behavior is an example [46].

Discussion

Principal Findings

Our results provide an initial basis for finding a consensus on the terminology of digital health promotion and prevention in settings through further research. The overall objective of this scoping review is to provide an overview of research targeting digital health promotion and primary prevention in settings. It focuses on assessing the range of scientific literature regarding various outcomes such as applied technology, targeted setting, and area of health promotion or prevention as well as on identifying research gaps that result from these findings. The results provide a first overview of scientific research on digital health promotion and prevention in settings. We have shown the wide range of applications of digital technologies in settings and identified research gaps. We systematically searched for empirical and nonempirical publications from 2010 to January 2020 and identified 200 references that met our inclusion criteria. The data analysis showed that there is a broad range and diversity of literature on digital health promotion and primary prevention in settings regarding various outcomes. Thus, we found publications focusing on a large spectrum of target groups, settings, and technologies as well as areas and methods of health promotion and prevention.

We discovered a wide range of analog, digital, and blended settings where digital health promotion and prevention interventions were applied. Regarding our results, it must be

considered that there is a scientific discourse about the definition and application of the concept *settings for health promotion and prevention*. Whereas some researchers support a broader approach, others recommend a narrower definition that only focuses on formal organizations [1]. As we have stated previously, we chose the broader approach.

To our knowledge, this is the first review focusing on digital health promotion and prevention in settings; therefore, we have to consider other studies and sensibly embed and discuss their results compared with our findings. Looking at digital interventions in the field of health promotion and prevention in general, one can see that settings are rarely taken into account. Thus, Norman et al [236] found in their systematic review that eHealth interventions for physical activity and nutrition were sometimes implemented in analog settings such as communities, schools, primary care, or workplaces. In addition, Bailey et al [237] identified in a scoping review that digital interventions for sexual health promotion were delivered in various analog settings, including schools, universities, and health care facilities; through web-based conditions; or through a combination of both, which supported the results of our scoping review.

There is a lack of a (uniform) definition of the terms *digital setting* and *blended setting*; therefore, a way has to be devised to bring together previous knowledge in this area and our results. In line with our presuppositions and results, Loss et al [47] concluded that social networking sites might be categorized as a setting for health promotion and prevention. Other than this publication, the existing scientific literature lacks an extensive and in-depth discussion about the existence and forms of digital (and blended) settings. What can be stated is that the most commonly found digital setting in this review, namely social media, is a widely discussed and used platform for health promotion and prevention interventions in terms of behavioral interventions. It is considered that social media interventions have the potential to facilitate access to health information as well as to expand the reach, improve the efficacy, and lower the costs of health promotion and prevention interventions [238-241]. However, it must be taken into account that the perspective of organizations (besides the workplace) has so far been largely absent from the scientific literature. The digitization of a formal organization in the field of health promotion and prevention is accompanied by massive structural changes. So far, however, these have not been scientifically investigated and evaluated—apart from digital health promotion in the workplace. For future research in the field of digital setting-based health promotion and prevention, this perspective must be taken into account.

What is most striking is the wide range of technologies used to implement health promotion and prevention in settings. These range from simple SMS text messages and smartphone apps to web-based programs offered on tablets in the waiting areas of health care providers. The existence of so many different technologies in the field of health promotion and prevention was identified in other reviews as well [34,242]. This diversity results from the rapid development of technologies in the health sector in the last few years, and it is reflected in the variety of terms that exist to define digitization in the health sector, for example, digital health, eHealth, mHealth, and Health 2.0. [16]. Another reasonable distinction of technology types comes from Lupton [238], who differentiates health promotion and prevention technologies between Web 1.0 and Web 2.0 applications. Web 2.0 refers to a newer generation of web-based applications that enable interaction and that allow users to generate, create, and share information [48]. This distinction can also be applied to the articles we identified.

Another central finding refers to the distinction between structural and behavioral health promotion and prevention in settings. Thus, most (170/200, 85%) of the publications on digital health promotion and prevention in settings can be categorized as behavioral health promotion and prevention in settings. In many cases, this means that the setting was *used* to reach a specific target group aiming at a change in their behavior. Consequently, we found a lack of articles concerning structural health promotion and prevention in settings, specifically a lack of interventions aiming at a change in setting-based and often organizational structures. It remains unclear how structural changes in settings can be achieved using digital interventions. Among the publications that were oriented toward structural health promotion and prevention in settings, only a few discussed structural changes in digital settings compared with changes in analog settings. This highlights that besides the lack of a scientific discussion on the existence and forms of digital settings for health promotion and prevention (which we have already pointed out), there is a research desideratum on how the structures of digital settings can be changed to improve the health of digital technology users.

Following the categorization of setting-based health promotion by Whitelaw et al [7], most of the digital health-promoting interventions can be classified under *passive model* where the setting is used as an access to individuals to try to get them to change their behavior. Interventions mostly miss out on taking into account the specific structures, social contexts, and interaction possibilities of the setting and cannot therefore be assigned to the other models. Nevertheless, we found a few examples in the literature where the digital setting-based intervention includes characteristics of the other intervention types. Table 1 presents examples of identified studies that were assigned to the types described by Whitelaw et al [7]. This overview demonstrates that existing theoretical frameworks from analog interventions can be used to structure digital health-promoting and preventive interventions in settings. Further research should focus on whether new types of setting-based health promotion emerge as a result of digitization.

Table 1. Digital interventions and types of setting-based health promotion (types based on Whitelaw et al [7], following Engelmann and Halkow [1]).

Type	Example of digital health promotion or prevention
Passive model • Individuals and their health-relevant behavior are at the center of interventions • Setting is used as an access to individuals and groups to provide them with the health-promoting intervention • Health promotion *in* the setting	**Bailey et al [46]** • The intervention consisted of a tablet that was displayed in a waiting room showing a website that aims to increase condom use and reduce sexually transmitted infections • The setting *health clinic* was used to access individuals to try to get them to change their health behavior • Setting or organizational structures were not considered
Active model • Focuses on the health behavior of individuals • Some behavior changes are inhibited or enabled by structural features of the setting • Structural modifications are required	**Bourdeaudhuij et al [49]** • An internet-based physical activity intervention was implemented in schools • The aim of the intervention was to change health behavior • A few structural features of the setting were considered, for example, technical equipment or environmental barriers
Vehicle model • Focuses on individual health issues • Projects are intended to initiate systemic changes • Projects serve as door openers for developments at other levels of the setting • The focus is on structural factors and their influence on health, even if individual health issues are addressed first	**Templeton et al [50]** • The aim was to create a web-based sexual health–promotion intervention to encourage young men in prisons to avail of regular sexual health checkups • The health intervention was used to achieve overarching goals, for example, to initiate collaborative partnerships to disrupt structures that lead to health inequities, to fulfill the human rights of prisoners, to develop a health-promoting prison
Organic model • Social systems are the product of the processes of individuals and groups living and acting together • Interventions bring systemic changes that affect the communication culture and community values in the setting • The setting is understood as the shaping framework for health, which is also shaped by setting members • Participation and empowerment of people to shape their environment are key concepts	**Sheats et al [43]** • Citizens use an app to collect data about their neighborhood that facilitate or hinder healthy living • They use the findings to advocate for change in partnership with local decision-makers and policy makers • Includes community-engaged methods to empower citizens • This transforms the possibilities of participation and communication in the community
Comprehensive model • Health is shaped by setting influences • Settings are understood as superordinate systems in which the individual has hardly any possibilities to shape them himself • Interventions are often implemented from the top down by the setting policy	**Delany et al [51]** • Web-based canteen ordering system in a school is altered with the methods of labeling, placement, and prompting to achieve healthier food choices of students • Change in the ordering system as a political decision of the school • Intervention is implemented from the top down; decision is not up to the students

We identified a wide spectrum of methods that are applied to achieve health promotion or prevention through digital interventions. It is noticeable that the methods of training and monitoring are mostly used for interventions on physical activity, whereas environmental change was mostly applied as a method within interventions on nutrition. A reason for that could be that there exist monitoring devices that are more effective or easier to use in the area of physical activity, for example, pedometers that automatically count steps compared with devices that monitor food intake. In addition, it is possible that environmental changes are easier implemented or more effective in the area of nutrition, for example, changes in the food environment at home compared with changes in physical activity options in settings. Comparing the methods used in the articles included in this review with those identified by Leppin [30] and Scherenberg [12], it is noticeable that normative and regulative proceedings as well as digital reminders were not applied. This could be because the development and introduction of new laws is difficult to realize through digital interventions and because digital reminders (eg, for vaccinations) are a method for individual health promotion rather than for setting-based interventions.

As we did not analyze the effectiveness of digital health promotion and prevention interventions in settings, this needs to be investigated in future research.

We found that most of the publications were empirical and the most used study design was an RCT. Only a small percentage had a qualitative design. Taking a closer look at the study designs, one can see that only a few studies followed a participatory design and that the RCTs usually analyzed the effectiveness of behavior-changing interventions. Therefore, one could conclude that research on digital health promotion and prevention in settings is at an advanced stage because digital apps are tested on their effectiveness rather than being in a previous stage of developing technologies or more fundamental

research. However, as we did not find many articles focusing on structural health promotion and prevention in settings, we cannot generalize this interpretation to the whole research field. As we have stated previously, there is a huge research gap concerning structural health promotion and prevention in settings, and research that is more fundamental is needed in this area. A possible explanation for the huge number of RCTs regarding behavioral health promotion and prevention and the lack of RCTs regarding structural health promotion and prevention could be that the effectiveness of changing health behavior (eg, physical activity) is more easily measurable than the effectiveness of changing structures or processes in formal and informal settings.

Limitations

This scoping review was conducted in compliance with the standards of the methodological approach outlined by Levac et al [28] and in accordance with the reporting guidelines of the PRISMA-ScR checklist [25]. Nevertheless, it includes some limitations that have to be taken into account.

First, the database search was very challenging because it was difficult to find suitable keywords that reflect the concept of a setting so as to be understood in the field of health promotion and prevention. Especially in combination with the term *technology*, the meaning of the word *setting* was not equivalent to our understanding of the concept. In addition, in none of the databases was the term *setting* well indexed. Another limitation derives from the fact that we searched for the terms *health promotion* and *health prevention* in general. We did not expand our search to specific health promotion areas such as physical activity or mental health. Therefore, we might have missed some articles if the terms *health promotion* or *health prevention* were not listed.

Second, our analysis and results are based on our presupposition that the definition of a setting for health promotion and prevention is applicable in the digital context too. Because of a lack of a definition of the term *digital setting*, the inclusion and exclusion criteria are based on the definitions of the term *setting* provided by Rosenbrock [3] and the WHO [4]. This presupposition has to be confirmed in the future to verify our results.

Third, we limited our search to the area of primary prevention and to health promotion and prevention in settings other than the workplace. These restrictions were chosen because of the huge breadth and diversity of the research area. It is possible that the inclusion of articles on secondary and tertiary prevention as well as articles regarding the workplace would have produced different results. In particular, more research has been conducted on digital health promotion in the workplace, and methods and technologies could potentially be applied to settings outside of the workplace.

Recommendations for Future Research

As we have stated previously, there are major research gaps in the research area of digital health promotion and prevention in settings, and there is a lack of a sufficient consideration of settings in the field of digital health. In future, the setting approach ought to be given greater importance, especially in the research area of digital health and digital health promotion and prevention. This includes developing definitions for the terms *digital setting* and *blended setting* and specifying their fields of application and the different forms that exist. In addition, key features of digital and blended settings have to be explored, especially with regard to their health impact. We currently know far too little about what elements define a digital setting. We need to further investigate the relevance of social interaction for the definition of digital settings.

Especially, there is a lack of research on structural health promotion and prevention in settings. As described previously, more research is needed on how digital technologies can change structures of formal and informal settings and how structures of digital settings themselves can be changed. Future research should focus on how structural change can be applied (eg, through organizational development), how it can be measured, and what stakeholders (eg, leaders or members of an organization) should be involved in this process. Therefore, a more detailed analysis of the identified articles on structural health promotion and prevention in settings seems useful.

It would also be desirable for future research to consider the development of digital health-promoting tools that focus on all target groups in a setting and that consequently consider the organization as a holistic setting with the different needs and requirements of the various target groups and thus do justice to the complexity and multidimensionality of the setting.

In our scoping review, we successfully mapped the research field of digital health promotion and prevention in settings regarding various outcomes. What remains unclear is the effectiveness of these interventions. Future scientific research should also examine the effects of digital interventions, especially the long-term effects. Digital health promotion and prevention technology can only establish itself when the benefits and risks for a population are proven. In addition, more participatory research is needed to ensure that the needs, wishes, (digital) competences, and possible fears of users are considered in the development, testing, and implementation of these technologies. It would also be interesting to see to what extent participatory research can take place in digital settings or to what extent the WHO's healthy setting principles (including community participation, partnership, empowerment, and equity [243]) can be achieved through digital tools. Such research would contribute to the acceptance and eventually to the extensive use of digital technologies for health promotion and prevention.

Conclusions

Our review provides an overview of the research on digital health promotion and prevention in settings. We demonstrated that the research field is complex, heterogeneous, and broad, and our findings reveal research gaps that urgently need to be addressed. A gap concerns study on structural health promotion and prevention in settings, and another concerns the definition of relevant terms, for example, *digital settings*. As the topic of digitization in setting-based health promotion and prevention is multidisciplinary, covering disciplines such as public health, organizational development, and the technology industry, in our opinion, the most important aspect is that different

disciplines combine their expertise and work collaboratively to bridge these research gaps. In this context, the participation of potential technology users is unavoidable to guarantee that the users' needs, especially the needs of the vulnerable, are considered. In addition, it is urgently necessary to clarify which significant improvements will result from the application of digital technologies. So far, there is insufficient basic research that provides information about which technologies in which health areas are best suited to certain settings or target groups. Therefore, in future studies, the great potential of digitization should be further explored based on validated scientific research. Especially, the benefits for the populations' health and the possibility of making health promotion and prevention interventions more effective and less expensive need to be analyzed.

Acknowledgments

The authors would like to thank the members of the scientific advisory board, Professor Dr Ralph Möhler (Bielefeld University), Professor Dr Kevin Dadaczynski (Fulda University of Applied Sciences), and Birgit Hollmann (Bielefeld University), who supported the review process through regular consultations and discussions. This scoping review was funded by the German private health insurance association, Verband der Privaten Krankenversicherung e.V., as part of a research project. The review was conducted independently by a research team at the University of Bielefeld. The authors would also like to acknowledge support for the article processing charge by the Deutsche Forschungsgemeinschaft and the Open Access Publication Fund of Bielefeld University.

Authors' Contributions

ALS and CG led the literature review and the research process, led critical reflections on the analysis and the concept, and prepared the manuscript, which was critically reviewed and revised by all authors. CD supervised the research, supported the writing process, and provided critical input. All authors have read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search strategy.
[PDF File (Adobe PDF File), 121 KB - jmir_v24i1e21063_app1.pdf]

Multimedia Appendix 2
Data extraction chart.
[PDF File (Adobe PDF File), 208 KB - jmir_v24i1e21063_app2.pdf]

Multimedia Appendix 3
Areas of health promotion and prevention (n=254 due to the fact that some publications report on more than one area).
[PNG File , 57 KB - jmir_v24i1e21063_app3.png]

References

1. Engelmann F, Halkow A. Der Setting-Ansatz in der Gesundheitsförderung: Genealogie, Konzeption, Praxis, Evidenzbasierung. Social Science Open Access Repository (SSOAR). 2008. URL: https://nbn-resolving.org/urn:nbn:de:0168-ssoar-294064 [accessed 2021-10-15]
2. Dadaczynski K, Baumgarten K, Hartmann T. Settingbasierte Gesundheitsförderung und Prävention. Präv Gesundheitsf 2016 Oct 7;11(4):214-221. [doi: 10.1007/s11553-016-0562-1]
3. Rosenbrock R. Prävention in Lebenswelten – der Setting-Ansatz. Z Allg Med 2015;91(5):213-219 [FREE Full text]
4. Health Promotion Glossary. World Health Organization. 1998. URL: https://www.who.int/healthpromotion/about/HPR%20Glossary%201998.pdf [accessed 2021-12-23]
5. Batras D, Duff C, Smith BJ. Organizational change theory: implications for health promotion practice. Health Promot Int 2016 Mar 14;31(1):231-241. [doi: 10.1093/heapro/dau098] [Medline: 25398838]
6. Seibt AC. Erklärungs- und Veränderungsmodelle II: Stufen und Phasen von Planungs- und Veränderungsprozessen. In: Blümel S, Franzkowiak P, Kaba-Schönstein L, editors. Leitbegriffe der Gesundheitsförderung und Prävention: Glossar zu Konzepten, Strategien und Methoden. Köln: Bundeszentrale für gesundheitliche Aufklärung; 2011:87-99.
7. Whitelaw S, Baxendale A, Bryce C, MacHardy L, Young I, Witney E. 'Settings' based health promotion: a review. Health Promot Int 2001 Dec;16(4):339-353. [doi: 10.1093/heapro/16.4.339] [Medline: 11733453]
8. Frees B, Koch W. ARD/ZDF-Onlinestudie 2018: Zuwachs bei medialer Internetnutzung und Kommunikation. Media Perspektiven 2018;9:398-413 [FREE Full text]

9. Baumann E, Czerwinski F. Erst mal Doktor Google fragen? Nutzung neuer Medien zur Information und zum Austausch über Gesundheitsthemen. In: Böcken J, Braun B, Meierjürgen R, editors. Gesundheitsmonitor 2015: Bürgerorientierung im Gesundheitswesen. Kooperationsprojekt der Bertelsmann Stiftung und der BARMER GEK. Gütersloh: Verlag Bertelsmann Stiftung; 2015:57-79.
10. Baumann E, Link E. Onlinebasierte Gesundheitskommunikation: Nutzung und Austausch von Gesundheitsinformationen über das Internet. In: Fischer F, Krämer A, editors. eHealth in Deutschland. Anforderungen und Potenziale innovativer Versorgungsstrukturen. Berlin, Heidelberg: Springer Vieweg; 2016:385-406.
11. Rossmann C, Stehr P. Gesundheitskommunikation im Internet. In: Schweiger W, Beck K, editors. Handbuch Online-Kommunikation. Wiesbaden: Springer VS; 2019:393-419.
12. Scherenberg V. Prävention Gesundheitsförderung in und mit digitalen Medien. In: Dockweiler C, Fischer F, editors. ePublic Health. Bern: Hogrefe; 2019:121-136.
13. Reinhardt K. Digitale Transformation der Organisation. Wiesbaden: Springer Fachmedien Wiesbaden; 2020.
14. Matusiewicz D, Pittelkau C, Elmer A, Addam M. Die digitale Transformation im Gesundheitswesen: Transformation, Innovation, Disruption. Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft; 2017.
15. Schott G, Hodgetts D. Health and digital gaming: the benefits of a community of practice. J Health Psychol 2006 Mar;11(2):309-316. [doi: 10.1177/1359105306061189] [Medline: 16464927]
16. Zeeb H, Pigeot I, Schüz B, Leibniz-WissenschaftsCampus Digital Public Health Bremen. [Digital public health-an overview]. Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Feb 09;63(2):137-144. [doi: 10.1007/s00103-019-03078-7] [Medline: 31919531]
17. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
18. Nievas Soriano BJ, Duarte S, Alonso A, Perales A, Carreño T. eHealth: advantages, disadvantages and guiding principles for the future. JMIR Preprints 2019:23. [doi: 10.2196/preprints.15366]
19. Ottawa Charter for Health Promotion. World Health Organization. 1986. URL: https://www.euro.who.int/__data/assets/pdf_file/0004/129532/Ottawa_Charter.pdf [accessed 2020-05-07]
20. Nittas V, Lun P, Ehrler F, Puhan MA, Mütsch M. Electronic Patient-Generated Health Data to Facilitate Disease Prevention and Health Promotion: Scoping Review. J Med Internet Res 2019 Oct 14;21(10):e13320 [FREE Full text] [doi: 10.2196/13320] [Medline: 31613225]
21. Alcântara CMD, Silva ANS, Pinheiro PNDC, Queiroz MVO. Digital technologies for promotion of healthy eating habits in teenagers. Rev Bras Enferm 2019 Apr;72(2):513-520 [FREE Full text] [doi: 10.1590/0034-7167-2018-0352] [Medline: 31017217]
22. Naslund JA, Aschbrenner KA, Araya R, Marsch LA, Unützer J, Patel V, et al. Digital technology for treating and preventing mental disorders in low-income and middle-income countries: a narrative review of the literature. Lancet Psychiatry 2017 Jun;4(6):486-500 [FREE Full text] [doi: 10.1016/S2215-0366(17)30096-2] [Medline: 28433615]
23. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
24. Fischer F. Digital interventions in prevention and health promotion: What kind of evidence do we have and what is needed? Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz 2020 Jun 30;63(6):674-680. [doi: 10.1007/s00103-020-03143-6] [Medline: 32355991]
25. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. Ann Intern Med 2018 Sep 04;169(7):467. [doi: 10.7326/M18-0850]
26. Peters MDJ, Godfrey CM, Khalil H, McInerney P, Parker D, Soares CB. Guidance for conducting systematic scoping reviews. Int J Evid Based Healthc 2015 Sep;13(3):141-146. [doi: 10.1097/XEB.0000000000000050] [Medline: 26134548]
27. Arksey H, O'Malley L. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology 2005 Feb;8(1):19-32. [doi: 10.1080/1364557032000119616]
28. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010 Sep 20;5(1):69 [FREE Full text] [doi: 10.1186/1748-5908-5-69] [Medline: 20854677]
29. Peters M, Godfrey C, McInerney P, Munn Z, Tricco A, Khalil H. Chapter 11: Scoping Reviews. In: Aromataris E, Munn Z, editors. Joanna Briggs Institute Reviewer's Manual. Adelaide: Joanna Briggs Institute; 2020.
30. Leppin A. Konzepte der Prävention, Strategien. In: Hurrelmann K, Klotz T, Haisch J, editors. Lehrbuch Prävention und Gesundheitsförderung. Bern: Verlag Hans Huber; 2014:36-44.
31. Altgeld T, Kolip P. Konzepte Strategien der Gesundheitsförderung. In: Hurrelmann K, Klotz T, Haisch J, editors. Lehrbuch Prävention und Gesundheitsförderung. Bern: Verlag Hans Huber; 2014:45-56.
32. GKV-Spitzenverband. Leitfaden Prävention—Handlungsfelder und Kriterien nach §; 2018 20 Abs 2 SGB V. 2020. URL: https://www.gkv-spitzenverband.de/media/dokumente/krankenversicherung_1/praevention__selbsthilfe__beratung/praevention/praevention_leitfaden/Leitfaden_Pravention_2020_barrierefrei.pdf [accessed 2022-01-04]
33. Evens A. Logic of the digital. London, UK: Bloomsbury; 2015.

34. O'Neil A, Cocker F, Rarau P, Baptista S, Cassimatis M, Barr Taylor C, et al. Using digital interventions to improve the cardiometabolic health of populations: a meta-review of reporting quality. J Am Med Inform Assoc 2017 Jul 01;24(4):867-879 [FREE Full text] [doi: 10.1093/jamia/ocw166] [Medline: 28339628]
35. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
36. Duncan DT, Aldstadt J, Whalen J, Melly SJ, Gortmaker SL. Validation of walk score for estimating neighborhood walkability: an analysis of four US metropolitan areas. Int J Environ Res Public Health 2011 Nov 04;8(11):4160-4179 [FREE Full text] [doi: 10.3390/ijerph8114160] [Medline: 22163200]
37. Anderson-Hanley C, Arciero PJ, Brickman AM, Nimon JP, Okuma N, Westen SC, et al. Exergaming and older adult cognition: a cluster randomized clinical trial. Am J Prev Med 2012 Feb;42(2):109-119. [doi: 10.1016/j.amepre.2011.10.016] [Medline: 22261206]
38. Dias JD, Mekaro MS, Cheng Lu JK, Otsuka JL, Fonseca LMM, Zem-Mascarenhas SH. Serious game development as a strategy for health promotion and tackling childhood obesity. Rev Lat Am Enfermagem 2016 Aug 15;24:e2759 [FREE Full text] [doi: 10.1590/1518-8345.1015.2759] [Medline: 27533268]
39. Luz S, Masoodian M, Cesario RR, Cesario M. Using a serious game to promote community-based awareness and prevention of neglected tropical diseases. Entertainment Computing 2016 Jun;15:43-55. [doi: 10.1016/j.entcom.2015.11.001]
40. Byun W, Lau E, Brusseau T. Feasibility and Effectiveness of a Wearable Technology-Based Physical Activity Intervention in Preschoolers: A Pilot Study. Int J Environ Res Public Health 2018 Aug 23;15(9):1821 [FREE Full text] [doi: 10.3390/ijerph15091821] [Medline: 30142911]
41. Duncan M, Staples V. The Impact of a School-Based Active Video Game Play Intervention on Children's Physical Activity During Recess. Human Movement 2010;11(1):95-99 [FREE Full text] [doi: 10.2478/v10038-009-0023-1]
42. Swindle TM, Ward WL, Whiteside-Mansell L. Facebook: The Use of Social Media to Engage Parents in a Preschool Obesity Prevention Curriculum. J Nutr Educ Behav 2018 Jan;50(1):4-10.e1 [FREE Full text] [doi: 10.1016/j.jneb.2017.05.344] [Medline: 29325661]
43. Sheats JL, Winter SJ, Romero PP, King AC. FEAST: Empowering Community Residents to Use Technology to Assess and Advocate for Healthy Food Environments. J Urban Health 2017 Apr 28;94(2):180-189 [FREE Full text] [doi: 10.1007/s11524-017-0141-6] [Medline: 28247054]
44. Yang Y, Kang B, Lee EY, Yang HK, Kim H, Lim S, et al. Effect of an obesity prevention program focused on motivating environments in childhood: a school-based prospective study. Int J Obes (Lond) 2017 Jul 20;41(7):1027-1034. [doi: 10.1038/ijo.2017.47] [Medline: 28216643]
45. King AC, Hekler EB, Grieco LA, Winter SJ, Sheats JL, Buman MP, et al. Effects of Three Motivationally Targeted Mobile Device Applications on Initial Physical Activity and Sedentary Behavior Change in Midlife and Older Adults: A Randomized Trial. PLoS One 2016;11(6):e0156370 [FREE Full text] [doi: 10.1371/journal.pone.0156370] [Medline: 27352250]
46. Bailey J, Tomlinson N, Hobbs L, Webster R. Challenges and opportunities in evaluating a digital sexual health intervention in a clinic setting: Staff and patient views. Digit Health 2017 Apr 11;3:2055207617704272 [FREE Full text] [doi: 10.1177/2055207617704272] [Medline: 29942593]
47. Loss J, Lindacher V, Curbach J. Online social networking sites-a novel setting for health promotion? Health Place 2014 Mar;26:161-170. [doi: 10.1016/j.healthplace.2013.12.012] [Medline: 24457613]
48. Kolt GS, Rosenkranz RR, Savage TN, Maeder AJ, Vandelanotte C, Duncan MJ, et al. WALK 2.0 - using Web 2.0 applications to promote health-related physical activity: a randomised controlled trial protocol. BMC Public Health 2013 May 03;13(1):436 [FREE Full text] [doi: 10.1186/1471-2458-13-436] [Medline: 23642010]
49. Bourdeaudhuij ID, Maes L, De Henauw S, De Vriendt T, Moreno LA, Kersting M, HELENA Study Group. Evaluation of a computer-tailored physical activity intervention in adolescents in six European countries: the Activ-O-Meter in the HELENA intervention study. J Adolesc Health 2010 May;46(5):458-466. [doi: 10.1016/j.jadohealth.2009.10.006] [Medline: 20413082]
50. Templeton M, Kelly C, Lohan M. Developing a Sexual Health Promotion Intervention With Young Men in Prisons: A Rights-Based Participatory Approach. JMIR Res Protoc 2019 Apr 29;8(4):e11829 [FREE Full text] [doi: 10.2196/11829] [Medline: 31033447]
51. Delaney T, Wyse R, Yoong SL, Sutherland R, Wiggers J, Ball K, et al. Cluster randomized controlled trial of a consumer behavior intervention to improve healthy food purchases from online canteens. Am J Clin Nutr 2017 Nov 27;106(5):1311-1320. [doi: 10.3945/ajcn.117.158329] [Medline: 28971849]
52. Allison S, Bauermeister JA, Bull S, Lightfoot M, Mustanski B, Shegog R, et al. The intersection of youth, technology, and new media with sexual health: moving the research agenda forward. J Adolesc Health 2012 Sep;51(3):207-212 [FREE Full text] [doi: 10.1016/j.jadohealth.2012.06.012] [Medline: 22921129]
53. Apolinário-Hagen J, Groenewold SD, Fritsche L, Kemper J, Krings L, Salewski C. Die Gesundheit Fernstudierender stärken. Präv Gesundheitsf 2017 Oct 17;13(2):151-158. [doi: 10.1007/s11553-017-0620-3]
54. Atkinson AM, Sumnall H, Measham F. Depictions of alcohol use in a UK Government partnered online social marketing campaign:. Drugs: Education, Prevention and Policy 2011 Jan 05;18(6):454-467. [doi: 10.3109/09687637.2010.534745]

55. Azevedo J, Padrão P, Gregório MJ, Almeida C, Moutinho N, Lien N, et al. A Web-Based Gamification Program to Improve Nutrition Literacy in Families of 3- to 5-Year-Old Children: The Nutriscience Project. J Nutr Educ Behav 2019 Mar;51(3):326-334. [doi: 10.1016/j.jneb.2018.10.008] [Medline: 30579894]
56. Backonja U, Hall AK, Thielke S. Older Adults' Current and Potential Uses of Information Technologies in a Changing World: A Theoretical Perspective. Int J Aging Hum Dev 2014 Dec 01;80(1):41-63 [FREE Full text] [doi: 10.1177/0091415015591109] [Medline: 26215298]
57. Bailey JV, Webster R, Hunter R, Griffin M, Freemantle N, Rait G, et al. The Men's Safer Sex project: intervention development and feasibility randomised controlled trial of an interactive digital intervention to increase condom use in men. Health Technol Assess 2016 Dec;20(91):1-124 [FREE Full text] [doi: 10.3310/hta20910] [Medline: 27966409]
58. Bamidis P, Konstantinidis E, Billis A, Frantzidis C, Tsolaki M, Tsolaki W, et al. A Web services-based exergaming platform for senior citizens: The long lasting memories project approach to e-health care. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2011 Presented at: Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 30 Aug.-3 Sept. 2011; Boston, MA, USA. [doi: 10.1109/iembs.2011.6090694]
59. Baranowski T, Blumberg F, Buday R, DeSmet A, Fiellin LE, Green CS, et al. Games for Health for Children-Current Status and Needed Research. Games Health J 2016 Feb;5(1):1-12 [FREE Full text] [doi: 10.1089/g4h.2015.0026] [Medline: 26262772]
60. Beck AM, Eyler AA, Aaron Hipp J, King AC, Tabak RG, Yan Y, et al. A multilevel approach for promoting physical activity in rural communities: a cluster randomized controlled trial. BMC Public Health 2019 Jan 30;19(1):126 [FREE Full text] [doi: 10.1186/s12889-019-6443-8] [Medline: 30700262]
61. Betsch C, Brewer NT, Brocard P, Davies P, Gaissmaier W, Haase N, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. Vaccine 2012 May 28;30(25):3727-3733. [doi: 10.1016/j.vaccine.2012.02.025] [Medline: 22365840]
62. Bickmore TW, Silliman RA, Nelson K, Cheng DM, Winter M, Henault L, et al. A randomized controlled trial of an automated exercise coach for older adults. J Am Geriatr Soc 2013 Oct;61(10):1676-1683. [doi: 10.1111/jgs.12449] [Medline: 24001030]
63. Brüll P, Ruiter RAC, Wiers RW, Kok G. Gaming for Safer Sex: Young German and Turkish People Report No Specific Culture-Related Preferences Toward Educational Games Promoting Safer Sex. Games Health J 2016 Sep 28. [doi: 10.1089/g4h.2016.0016] [Medline: 27680494]
64. Buendía Eisman A, Arias Santiago S, Moreno-Gimenez J, Cabrera-León A, Prieto L, Castillejo I, et al. An Internet-based programme to promote adequate UV exposure behaviour in adolescents in Spain. J Eur Acad Dermatol Venereol 2013 Apr;27(4):442-453. [doi: 10.1111/j.1468-3083.2012.04455.x] [Medline: 22329864]
65. Bull SS, Levine DK, Black SR, Schmiege SJ, Santelli J. Social media-delivered sexual health intervention: a cluster randomized controlled trial. Am J Prev Med 2012 Nov;43(5):467-474 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.022] [Medline: 23079168]
66. Buller DB, Andersen PA, Walkosz BJ, Scott MD, Beck L, Cutter GR. Effect of an intervention on observed sun protection by vacationers in a randomized controlled trial at North American resorts. Prev Med 2017 Jun;99:29-36 [FREE Full text] [doi: 10.1016/j.ypmed.2017.01.014] [Medline: 28189810]
67. Buman MP, Winter SJ, Sheats JL, Hekler EB, Otten JJ, Grieco LA, et al. The Stanford Healthy Neighborhood Discovery Tool: a computerized tool to assess active living environments. Am J Prev Med 2013 Apr;44(4):e41-e47 [FREE Full text] [doi: 10.1016/j.amepre.2012.11.028] [Medline: 23498112]
68. Byrd-Bredbenner C, Delaney C, Martin-Biggers J, Koenings M, Quick V. The marketing plan and outcome indicators for recruiting and retaining parents in the HomeStyles randomized controlled trial. Trials 2017 Nov 15;18(1):540 [FREE Full text] [doi: 10.1186/s13063-017-2262-3] [Medline: 29141692]
69. Byrd-Bredbenner C, Martin-Biggers J, Koenings M, Quick V, Hongu N, Worobey J. HomeStyles, A Web-Based Childhood Obesity Prevention Program for Families With Preschool Children: Protocol for a Randomized Controlled Trial. JMIR Res Protoc 2017 Apr 25;6(4):e73 [FREE Full text] [doi: 10.2196/resprot.7544] [Medline: 28442452]
70. Byrd-Bredbenner C, Martin-Biggers J, Povis GA, Worobey J, Hongu N, Quick V. Promoting healthy home environments and lifestyles in families with preschool children: HomeStyles, a randomized controlled trial. Contemp Clin Trials 2018 Jan;64:139-151. [doi: 10.1016/j.cct.2017.10.012] [Medline: 29079392]
71. Calear AL, Batterham PJ, Poyser CT, Mackinnon AJ, Griffiths KM, Christensen H. Cluster randomised controlled trial of the e-couch Anxiety and Worry program in schools. J Affect Disord 2016 May 15;196:210-217. [doi: 10.1016/j.jad.2016.02.049] [Medline: 26926660]
72. Calear AL, Christensen H, Griffiths KM, Mackinnon A. The Y-Worri Project: study protocol for a randomised controlled trial. Trials 2013 Mar 19;14:76 [FREE Full text] [doi: 10.1186/1745-6215-14-76] [Medline: 23506049]
73. Caperchione CM, Kolt GS, Savage TN, Rosenkranz RR, Maeder AJ, Vandelanotte C, et al. WALK 2.0: examining the effectiveness of Web 2.0 features to increase physical activity in a 'real world' setting: an ecological trial. BMJ Open 2014 Oct 10;4(10):e006374 [FREE Full text] [doi: 10.1136/bmjopen-2014-006374] [Medline: 25304191]

74. Carah N, Meurk C, Angus D. Online self-expression and experimentation as 'reflectivism': Using text analytics to examine the participatory forum Hello Sunday Morning. Health (London) 2017 Mar 26;21(2):119-135. [doi: 10.1177/1363459315596799] [Medline: 26216897]
75. Carlfjord S, Lindberg M, Andersson A. Staff perceptions of addressing lifestyle in primary health care: a qualitative evaluation 2 years after the introduction of a lifestyle intervention tool. BMC Fam Pract 2012 Oct 10;13(1):99 [FREE Full text] [doi: 10.1186/1471-2296-13-99] [Medline: 23052150]
76. Carr LJ, Dunsiger SI, Marcus BH. Walk score™ as a global estimate of neighborhood walkability. Am J Prev Med 2010 Nov;39(5):460-463 [FREE Full text] [doi: 10.1016/j.amepre.2010.07.007] [Medline: 20965384]
77. Carr LJ, Dunsiger SI, Lewis B, Ciccolo JT, Hartman S, Bock B, et al. Randomized controlled trial testing an internet physical activity intervention for sedentary adults. Health Psychol 2013 Mar;32(3):328-336 [FREE Full text] [doi: 10.1037/a0028962] [Medline: 22823069]
78. Cavallo DN, Brown JD, Tate DF, DeVellis RF, Zimmer C, Ammerman AS. The role of companionship, esteem, and informational support in explaining physical activity among young women in an online social network intervention. J Behav Med 2014 Oct 1;37(5):955-966 [FREE Full text] [doi: 10.1007/s10865-013-9534-5] [Medline: 24081454]
79. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
80. Champion KE, Newton NC, Stapinski L, Slade T, Barrett EL, Teesson M. A cross-validation trial of an Internet-based prevention program for alcohol and cannabis: Preliminary results from a cluster randomised controlled trial. Aust N Z J Psychiatry 2016 Jan;50(1):64-73. [doi: 10.1177/0004867415577435] [Medline: 25801662]
81. Champion KE, Newton NC, Stapinski LA, Teesson M. Effectiveness of a universal internet-based prevention program for ecstasy and new psychoactive substances: a cluster randomized controlled trial. Addiction 2016 Aug 13;111(8):1396-1405. [doi: 10.1111/add.13345] [Medline: 26880476]
82. Champion KE, Newton NC, Stapinski L, Teesson M. Cluster randomised controlled trial of an online intervention to prevent ecstasy and new psychoactive substance use among adolescents: final results and implications for implementation. BMJ Open 2018 Nov 25;8(11):e020433 [FREE Full text] [doi: 10.1136/bmjopen-2017-020433] [Medline: 30478103]
83. Chandra PS, Sowmya H, Mehrotra S, Duggal M. 'SMS' for mental health—feasibility and acceptability of using text messages for mental health promotion among young women from urban low income settings in India. Asian J Psychiatr 2014 Oct;11:59-64. [doi: 10.1016/j.ajp.2014.06.008] [Medline: 25453699]
84. Chen ACC, Amresh A. Developing a Bilingual, Computer-Tailored, HPV Vaccination Promotion Intervention Targeting Latino Parents. In: Proceedings of the 5th International Conference on Digital Health 2015. 2015 Presented at: DH '15: Digital Health 2015 Conference; May 18 - 20, 201; Florence Italy p. 59-64. [doi: 10.1145/2750511.2750522]
85. Chen L, Wang W, Du X, Rao X, van Velthoven MH, Yang R, et al. Effectiveness of a smart phone app on improving immunization of children in rural Sichuan Province, China: study protocol for a paired cluster randomized controlled trial. BMC Public Health 2014 Mar 20;14(1):262 [FREE Full text] [doi: 10.1186/1471-2458-14-262] [Medline: 24645829]
86. Cook TL, De Bourdeaudhuij I, Maes L, Haerens L, Grammatikaki E, Widhalm K, et al. Moderators of the effectiveness of a web-based tailored intervention promoting physical activity in adolescents: the HELENA Activ-O-Meter. J Sch Health 2014 Apr 11;84(4):256-266. [doi: 10.1111/josh.12140] [Medline: 24617909]
87. Cook T, De Bourdeaudhuij I, Maes L, Haerens L, Grammatikaki E, Widhalm K, et al. Psychosocial determinants and perceived environmental barriers as mediators of the effectiveness of a web-based tailored intervention promoting physical activity in adolescents: the HELENA Activ-O-Meter. J Phys Act Health 2014 May;11(4):741-751. [doi: 10.1123/jpah.2012-0120] [Medline: 23676337]
88. Crawford G, Maycock B, Tobin R, Brown G, Lobo R. Prevention of HIV and Other Sexually Transmissible Infections in Expatriates and Traveler Networks: Qualitative Study of Peer Interaction in an Online Forum. J Med Internet Res 2018 Sep 05;20(9):e10787 [FREE Full text] [doi: 10.2196/10787] [Medline: 30185404]
89. Cullen KW, Thompson D, Chen T. Outcome Evaluation of Family Eats. Health Educ Behav 2017 Feb 09;44(1):32-40. [doi: 10.1177/1090198116643917] [Medline: 27198535]
90. „fit4future“: 1.200 neue Schulen und Kitas. DAK Gesundheit. 2020. URL: https://www.dak.de/dak/bundesthemen/fit4future-1-200-neue-schulen-und-kitas-2103614.html [accessed 2020-05-15]
91. Daley MF, Narwaney KJ, Shoup JA, Wagner NM, Glanz JM. Addressing Parents' Vaccine Concerns: A Randomized Trial of a Social Media Intervention. Am J Prev Med 2018 Jul;55(1):44-54. [doi: 10.1016/j.amepre.2018.04.010] [Medline: 29773490]
92. Davies C, Corry K, Van IA, Vandelanotte C, Caperchione C, Mummery WK. Prospective associations between intervention components and website engagement in a publicly available physical activity website: the case of 10,000 Steps Australia. J Med Internet Res 2012 Jan 11;14(1):e4 [FREE Full text] [doi: 10.2196/jmir.1792] [Medline: 22260810]
93. Degroote L, Plaete J, De Bourdeaudhuij I, Verloigne M, Van Stappen V, De Meester A, et al. The Effect of the eHealth Intervention 'MyPlan 1.0' on Physical Activity in Adults Who Visit General Practice: A Quasi-Experimental Trial. Int J Environ Res Public Health 2018 Jan 30;15(2):228 [FREE Full text] [doi: 10.3390/ijerph15020228] [Medline: 29385770]

94. Dixon BE, Zimet GD, Xiao S, Tu W, Lindsay B, Church A, et al. An Educational Intervention to Improve HPV Vaccination: A Cluster Randomized Trial. Pediatrics 2019 Jan 10;143(1):e20181457. [doi: 10.1542/peds.2018-1457] [Medline: 30530637]
95. Dour CA, Horacek TM, Schembre SM, Lohse B, Hoerr S, Kattelmann K, et al. Process evaluation of Project WebHealth: a nondieting Web-based intervention for obesity prevention in college students. J Nutr Educ Behav 2013 Jul;45(4):288-295. [doi: 10.1016/j.jneb.2012.10.001] [Medline: 23410994]
96. Duncan DT. What's your Walk Score®?: Web-based neighborhood walkability assessment for health promotion and disease prevention. Am J Prev Med 2013 Aug;45(2):244-245. [doi: 10.1016/j.amepre.2013.04.008] [Medline: 23867033]
97. Dway NS, Soonthornworasiri N, Jandee K, Lawpoolsri S, Pan-Ngum W, Sinthuvanich D, et al. Effects of edutainment on knowledge and perceptions of Lisu mothers about the immunisation of their children. Health Education Journal 2015 Jan 23;75(2):131-143. [doi: 10.1177/0017896915569086]
98. Ehlers DK, Huberty JL, de Vreede GJ. Can an evidence-based book club intervention delivered via a tablet computer improve physical activity in middle-aged women? Telemed J E Health 2015 Feb;21(2):125-131. [doi: 10.1089/tmj.2013.0360] [Medline: 25526014]
99. Ekberg J, Timpka T, Angbratt M, Frank L, Norén AM, Hedin L, et al. Design of an online health-promoting community: negotiating user community needs with public health goals and service capabilities. BMC Health Serv Res 2013 Jul 04;13(1):258 [FREE Full text] [doi: 10.1186/1472-6963-13-258] [Medline: 23826944]
100. Ekram S, Debiec KE, Pumper MA, Moreno MA. Content and Commentary: HPV Vaccine and YouTube. J Pediatr Adolesc Gynecol 2019 Apr;32(2):153-157. [doi: 10.1016/j.jpag.2018.11.001] [Medline: 30445163]
101. Eschenbeck H, Lehner L, Hofmann H, Bauer S, Becker K, Diestelkamp S, ProHEAD Consortium. School-based mental health promotion in children and adolescents with StresSOS using online or face-to-face interventions: study protocol for a randomized controlled trial within the ProHEAD Consortium. Trials 2019 Jan 18;20(1):64 [FREE Full text] [doi: 10.1186/s13063-018-3159-5] [Medline: 30658675]
102. Ezendam NP, Burg J, Borsboom G, van Empelen P, Oenema A. Differential effects of the computer-tailored FATaintPHAT programme on dietary behaviours according to sociodemographic, cognitive and home environmental factors. Public Health Nutr 2012 Dec 21;17(02):431-439. [doi: 10.1017/s1368980012005344]
103. Ezendam NPM, Brug J, Oenema A. Evaluation of the Web-based computer-tailored FATaintPHAT intervention to promote energy balance among adolescents: results from a school cluster randomized trial. Arch Pediatr Adolesc Med 2012 Mar 05;166(3):248-255. [doi: 10.1001/archpediatrics.2011.204] [Medline: 22064878]
104. Fiellin LE, Kyriakides TC, Hieftje KD, Pendergrass TM, Duncan LR, Dziura JD, et al. The design and implementation of a randomized controlled trial of a risk reduction and human immunodeficiency virus prevention videogame intervention in minority adolescents: PlayForward: Elm City Stories. Clin Trials 2016 Aug;13(4):400-408. [doi: 10.1177/1740774516637871] [Medline: 27013483]
105. Gabarron E, Serrano JA, Wynn R, Armayones M. Avatars using computer/smartphone mediated communication and social networking in prevention of sexually transmitted diseases among North-Norwegian youngsters. BMC Med Inform Decis Mak 2012 Oct 30;12(1):120 [FREE Full text] [doi: 10.1186/1472-6947-12-120] [Medline: 23110684]
106. Ganz T, Braun M, Laging M, Schermelleh-Engel K, Michalak J, Heidenreich T. Effects of a stand-alone web-based electronic screening and brief intervention targeting alcohol use in university students of legal drinking age: A randomized controlled trial. Addict Behav 2018 Feb;77:81-88. [doi: 10.1016/j.addbeh.2017.09.017] [Medline: 28985586]
107. García-Camacha A, García-Camacha I, Martínez-Andrés M, Notario-Pacheco B, Rodríguez-Martín B. Pilot testing the effectiveness of the Healthy Ageing Supported by Internet and Community programme for promoting healthy lifestyles for people over 65 years of age. Scand J Caring Sci 2020 Sep 15;34(3):636-647. [doi: 10.1111/scs.12765] [Medline: 31614020]
108. Gilliland J, Sadler R, Clark A, O'Connor C, Milczarek M, Doherty S. Using a Smartphone Application to Promote Healthy Dietary Behaviours and Local Food Consumption. Biomed Res Int 2015;2015:841368 [FREE Full text] [doi: 10.1155/2015/841368] [Medline: 26380298]
109. Gold J, Pedrana AE, Stoove MA, Chang S, Howard S, Asselin J, et al. Developing health promotion interventions on social networking sites: recommendations from The FaceSpace Project. J Med Internet Res 2012 Feb 28;14(1):e30 [FREE Full text] [doi: 10.2196/jmir.1875] [Medline: 22374589]
110. González C, Herrero P, Cubero JM, Iniesta JM, Hernando ME, García-Sáez G, et al. PREDIRCAM eHealth platform for individualized telemedical assistance for lifestyle modification in the treatment of obesity, diabetes, and cardiometabolic risk prevention: a pilot study (PREDIRCAM 1). J Diabetes Sci Technol 2013 Jul 01;7(4):888-897 [FREE Full text] [doi: 10.1177/193229681300700411] [Medline: 23911170]
111. Grabowski D. Health identity, participation and knowledge: A qualitative study of a computer game for health education among adolescents in Denmark. Health Education Journal 2013 Jan 14;72(6):761-768. [doi: 10.1177/0017896912469559]
112. Graham ML, Uesugi KH, Niederdeppe J, Gay GK, Olson CM. The theory, development, and implementation of an e-intervention to prevent excessive gestational weight gain: e-Moms Roc. Telemed J E Health 2014 Dec;20(12):1135-1142 [FREE Full text] [doi: 10.1089/tmj.2013.0354] [Medline: 25354350]

113. Graham ML, Strawderman MS, Demment M, Olson CM. Does Usage of an eHealth Intervention Reduce the Risk of Excessive Gestational Weight Gain? Secondary Analysis From a Randomized Controlled Trial. J Med Internet Res 2017 Jan 09;19(1):e6 [FREE Full text] [doi: 10.2196/jmir.6644] [Medline: 28069560]
114. Greaney ML, Puleo E, Bennett GG, Haines J, Viswanath K, Gillman MW, et al. Factors associated with choice of web or print intervention materials in the healthy directions 2 study. Health Educ Behav 2014 Feb;41(1):52-62 [FREE Full text] [doi: 10.1177/1090198113486803] [Medline: 23720532]
115. Grim M, Hortz B, Petosa R. Impact Evaluation of a Pilot Web-Based Intervention to Increase Physical Activity. Am J Health Promot 2011 Mar 01;25(4):227-230. [doi: 10.4278/ajhp.081216-arb-307]
116. Grindell C, Mawson S, Gerrish K, Parker S, Bissell P. Exploring the acceptability and usability of a novel social innovation to encourage physical activity: The iStep prototype. Health Soc Care Community 2019 Mar 26;27(2):383-391. [doi: 10.1111/hsc.12656] [Medline: 30255638]
117. Gubrium AC, Fiddian-Green A, Lowe S, DiFulvio G, Del Toro-Mejías L. Measuring Down: Evaluating Digital Storytelling as a Process for Narrative Health Promotion. Qual Health Res 2016 Nov 09;26(13):1787-1801. [doi: 10.1177/1049732316649353] [Medline: 27184518]
118. Gustafson A, Jilcott Pitts S, McQuerry K, Babtunde O, Mullins J. A Mentor-Led Text-Messaging Intervention Increases Intake of Fruits and Vegetables and Goal Setting for Healthier Dietary Consumption among Rural Adolescents in Kentucky and North Carolina, 2017. Nutrients 2019 Mar 11;11(3):593 [FREE Full text] [doi: 10.3390/nu11030593] [Medline: 30862118]
119. Haruna H, Hu X, Chu S, Mellecker R. Initial Validation of the MAKE Framework: A Comprehensive Instrument for Evaluating the Efficacy of Game-Based Learning and Gamification in Adolescent Sexual Health Literacy. Ann Glob Health 2019 Feb 28;85(1):A [FREE Full text] [doi: 10.5334/aogh.1110] [Medline: 30873788]
120. Haruna H, Zainuddin Z, Mellecker RR, Chu SK, Hu X. An iterative process for developing digital gamified sexual health education for adolescent students in low-tech settings. ILS 2019 Nov 11;120(11/12):723-742. [doi: 10.1108/ils-07-2019-0066]
121. Haug S, Paz Castro R, Meyer C, Filler A, Kowatsch T, Schaub MP. A Mobile Phone-Based Life Skills Training Program for Substance Use Prevention Among Adolescents: Pre-Post Study on the Acceptance and Potential Effectiveness of the Program, Ready4life. JMIR Mhealth Uhealth 2017;5(10):e143 [FREE Full text] [doi: 10.2196/mhealth.8474] [Medline: 28978498]
122. Haug S, Paz Castro R, Wenger A, Schaub MP. Efficacy of a mobile phone-based life-skills training program for substance use prevention among adolescents: study protocol of a cluster-randomised controlled trial. BMC Public Health 2018 Sep 10;18(1):1102 [FREE Full text] [doi: 10.1186/s12889-018-5969-5] [Medline: 30200928]
123. Helander E, Kaipainen K, Korhonen I, Wansink B. Factors related to sustained use of a free mobile app for dietary self-monitoring with photography and peer feedback: retrospective cohort study. J Med Internet Res 2014;16(4):e109 [FREE Full text] [doi: 10.2196/jmir.3084] [Medline: 24735567]
124. Horstmann D, Tolks D, Dadaczynski K, Paulus P. Förderung des Wohlbefindens durch „Gamification“. Präv Gesundheitsf 2018 Jul 20;13(4):305-311. [doi: 10.1007/s11553-018-0659-9]
125. Huang S, Hung W, Shyu M, Chang K, Chen C. Web-based intervention to promote physical activity in Taiwanese children. J Pediatr Nurs 2019 Mar;45:e35-e43. [doi: 10.1016/j.pedn.2018.12.016] [Medline: 30700375]
126. Hyden C, Cohall A. Innovative approaches to using new media and technology in health promotion for adolescents and young adults. In: AM:STARs: Advances In Health Promotion for Adolescents and Young Adults. Itasca, Illinois, United States: American Academy of Pediatrics; 2005:498-520.
127. Irwin B, Kurz D, Chalin P, Thompson N. Testing the Efficacy of OurSpace, a Brief, Group Dynamics-Based Physical Activity Intervention: A Randomized Controlled Trial. J Med Internet Res 2016 May 06;18(4):e87 [FREE Full text] [doi: 10.2196/jmir.5342] [Medline: 27154301]
128. Jaganath D, Gill HK, Cohen AC, Young SD. Harnessing Online Peer Education (HOPE): integrating C-POL and social media to train peer leaders in HIV prevention. AIDS Care 2012 May;24(5):593-600 [FREE Full text] [doi: 10.1080/09540121.2011.630355] [Medline: 22149081]
129. Kattelmann KK, Bredbenner CB, White AA, Greene GW, Hoerr SL, Kidd T, et al. The effects of Young Adults Eating and Active for Health (YEAH): a theory-based Web-delivered intervention. J Nutr Educ Behav 2014 Nov;46(6):S27-S41. [doi: 10.1016/j.jneb.2014.08.007] [Medline: 25457733]
130. Kattelmann KK, White AA, Greene GW, Byrd-Bredbenner C, Hoerr SL, Horacek TM, et al. Development of Young Adults Eating and Active for Health (YEAH) internet-based intervention via a community-based participatory research model. J Nutr Educ Behav 2014 Mar;46(2):S10-S25. [doi: 10.1016/j.jneb.2013.11.006] [Medline: 24456834]
131. InCogito. Kaufmännische Krankenkasse. 2020. URL: https://www.kkh.de/leistungen/praevention-vorsorge/gesundheitsfoerderung-setting/incogito [accessed 2020-05-15]
132. Kim HK, Niederdeppe J, Guillory J, Graham M, Olson C, Gay G. Determinants of Pregnant Women's Online Self-Regulatory Activities for Appropriate Gestational Weight Gain. Health Commun 2015 Sep 10;30(9):922-932 [FREE Full text] [doi: 10.1080/10410236.2014.905900] [Medline: 25205417]

133. King A, Winter S, Sheats J, Rosas LG, Buman MP, Salvo D, et al. Leveraging Citizen Science and Information Technology for Population Physical Activity Promotion. Transl J Am Coll Sports Med 2016 May 15;1(4):30-44 [FREE Full text] [Medline: 27525309]
134. Knöll M, Moar M. The Space of Digital Health Games. International Journal of Computer Science in Sport 2012;11(1):61-72 [FREE Full text]
135. Knöll M. Mobile Partizipation in der gesundheitsfördernden Stadtgestaltung—zwei Fallbeispiele zu Datenerfassung und Interaktion im Stadtraum. In: Baumgart S, Köckler H, Ritzinger A, Rüdiger A, editors. Planung für gesundheitsfördernde Städte. Hannover: ARL – Akademie für Raumentwicklung in der Leibniz-Gemeinschaft; 2018:387-401.
136. Knöll M. Urban health games. Collaborative, expressive and reflective. Dissertation, University of Stuttgart. 2012. URL: https://elib.uni-stuttgart.de/handle/11682/109 [accessed 2021-12-23]
137. Knowlden AP, Conrad E. Two-Year Outcomes of the Enabling Mothers to Prevent Pediatric Obesity Through Web-Based Education and Reciprocal Determinism (EMPOWER) Randomized Control Trial. Health Educ Behav 2018 Apr 27;45(2):262-276. [doi: 10.1177/1090198117732604] [Medline: 28954544]
138. Knowlden AP, Sharma M, Cottrell RR, Wilson BRA, Johnson ML. Impact evaluation of Enabling Mothers to Prevent Pediatric Obesity through Web-Based Education and Reciprocal Determinism (EMPOWER) Randomized Control Trial. Health Educ Behav 2015 Apr;42(2):171-184. [doi: 10.1177/1090198114547816] [Medline: 25161168]
139. Konstantinidis E, Billis A, Hlauschek W, Panek P, Bamidis P. Integration of cognitive and physical training in a smart home environment for the elderly people. Studies in health technology and informatics 2010;160(1):58-62. [doi: 10.3233/978-1-60750-588-4-58]
140. Kousoulis AA, Kympouropoulos SP, Pouli DK, Economopoulos KP, Vardavas CI. From the Classroom to Facebook: A Fresh Approach for Youth Tobacco Prevention. Am J Health Promot 2016 May;30(5):390-393. [doi: 10.1177/0890117116646345] [Medline: 27404648]
141. Kröninger-Jungaberle H, Nagy E, von Heyden M, DuBois F, Ullrich J, Wippermann C, et al. REBOUND: A media-based life skills and risk education programme. Health Education Journal 2014 Nov 16;74(6):705-719. [doi: 10.1177/0017896914557097]
142. Kuosmanen T, Fleming T, Barry M. The implementation of SPARX-R computerized mental health program in alternative education: Exploring the factors contributing to engagement and dropout. Children and Youth Services Review 2018 Jan;84:176-184. [doi: 10.1016/j.childyouth.2017.11.032]
143. Kuosmanen T, Fleming TM, Barry MM. Using Computerized Mental Health Programs in Alternative Education: Understanding the Requirements of Students and Staff. Health Commun 2018 Jun 20;33(6):753-761. [doi: 10.1080/10410236.2017.1309620] [Medline: 28426253]
144. Lakerveld J, Mackenbach JD, de Boer F, Brandhorst B, Broerse JEW, de Bruijn G, et al. Improving cardiometabolic health through nudging dietary behaviours and physical activity in low SES adults: design of the Supreme Nudge project. BMC Public Health 2018 Jul 20;18(1):899 [FREE Full text] [doi: 10.1186/s12889-018-5839-1] [Medline: 30029600]
145. Lana A, del Valle MO, López S, Faya-Ornia G, López ML. Study protocol of a randomized controlled trial to improve cancer prevention behaviors in adolescents and adults using a web-based intervention supplemented with SMS. BMC Public Health 2013 Apr 17;13(1):357 [FREE Full text] [doi: 10.1186/1471-2458-13-357] [Medline: 23594381]
146. Larsen B, Marcus B, Pekmezi D, Hartman S, Gilmer T. A Web-Based Physical Activity Intervention for Spanish-Speaking Latinas: A Costs and Cost-Effectiveness Analysis. J Med Internet Res 2017 Feb 22;19(2):e43 [FREE Full text] [doi: 10.2196/jmir.6257] [Medline: 28228368]
147. Laska MN, Sevcik SM, Moe SG, Petrich CA, Nanney MS, Linde JA, et al. A 2-year young adult obesity prevention trial in the US: Process evaluation results. Health Promot Int 2016 Dec 30;31(4):793-800 [FREE Full text] [doi: 10.1093/heapro/dav066] [Medline: 26135586]
148. Lima T, Beltrame J, Niquini C, Barbosa B, Davis C. Design of a Mixed-Reality Serious Game to Tackle a Public Health Problem. In: Proceedings of the 17th International Conference on Entertainment Computing (ICEC). 2018 Presented at: 17th International Conference on Entertainment Computing (ICEC); Sep 2018; Poznan, Poland p. 305-309 URL: https://hal.inria.fr/hal-02128602/document
149. Little P, Stuart B, Hobbs FDR, Moore M, Barnett J, Popoola D, et al. An internet-delivered handwashing intervention to modify influenza-like illness and respiratory infection transmission (PRIMIT): a primary care randomised trial. The Lancet 2015 Oct;386(10004):1631-1639. [doi: 10.1016/s0140-6736(15)60127-1]
150. Loh IH, Schwendler T, Trude AC, Anderson Steeves ET, Cheskin LJ, Lange S, et al. Implementation of Text-Messaging and Social Media Strategies in a Multilevel Childhood Obesity Prevention Intervention: Process Evaluation Results. Inquiry 2018 Jun 04;55:46958018779189 [FREE Full text] [doi: 10.1177/0046958018779189] [Medline: 29865969]
151. Lombard C, Brennan L, Reid M, Klassen KM, Palermo C, Walker T, et al. Communicating health-Optimising young adults' engagement with health messages using social media: Study protocol. Nutr Diet 2018 Nov 15;75(5):509-519. [doi: 10.1111/1747-0080.12448] [Medline: 30009396]
152. Lumsden C, Wolf R, Contento I, Basch C, Zybert P, Koch P, et al. Feasibility, Acceptability, and Short-term Behavioral Impact of the MySmileBuddy Intervention for Early Childhood Caries. J Health Care Poor Underserved 2019;30(1):59-69. [doi: 10.1353/hpu.2019.0007] [Medline: 30827969]

153. Lyson HC, Le GM, Zhang J, Rivadeneira N, Lyles C, Radcliffe K, et al. Social Media as a Tool to Promote Health Awareness: Results from an Online Cervical Cancer Prevention Study. J Cancer Educ 2019 Aug 11;34(4):819-822 [FREE Full text] [doi: 10.1007/s13187-018-1379-8] [Medline: 29948924]
154. Maes L, Cook TL, Ottovaere C, Matthijs C, Moreno LA, Kersting M, et al. Pilot evaluation of the HELENA (Healthy Lifestyle in Europe by Nutrition in Adolescence) Food-O-Meter, a computer-tailored nutrition advice for adolescents: a study in six European cities. Public Health Nutr 2011 Jul;14(7):1292-1302. [doi: 10.1017/S1368980010003563] [Medline: 21281542]
155. Magoc D, Tomaka J, Bridges-Arzaga A. Using the web to increase physical activity in college students. Am J Health Behav 2011;35(2):142-154. [doi: 10.5993/ajhb.35.2.2] [Medline: 21204677]
156. Maher C, Ferguson M, Vandelanotte C, Plotnikoff R, De Bourdeaudhuij I, Thomas S, et al. A Web-Based, Social Networking Physical Activity Intervention for Insufficiently Active Adults Delivered via Facebook App: Randomized Controlled Trial. J Med Internet Res 2015 Jul;17(7):e174 [FREE Full text] [doi: 10.2196/jmir.4086] [Medline: 26169067]
157. Mahmud AJ, Olander E, Eriksén S, Haglund BJ. Health communication in primary health care -a case study of ICT development for health promotion. BMC Med Inform Decis Mak 2013 Jan 30;13:17 [FREE Full text] [doi: 10.1186/1472-6947-13-17] [Medline: 23363566]
158. Malmberg M, Kleinjan M, Overbeek G, Vermulst A, Monshouwer K, Lammers J, et al. Effectiveness of the 'Healthy School and Drugs' prevention programme on adolescents' substance use: a randomized clustered trial. Addiction 2014 Jun;109(6):1031-1040. [doi: 10.1111/add.12526] [Medline: 24612164]
159. Manning J, Vandeusen K. Suicide prevention in the dot com era: technological aspects of a university suicide prevention program. J Am Coll Health 2011 Apr 13;59(5):431-433. [doi: 10.1080/07448480903540507] [Medline: 21500064]
160. Manzoor A, Mollee J, Araujo E, van Halteren A, Klein M. Online Sharing of Physical Activity: Does It Accelerate the Impact of a Health Promotion Program? In: Proceedings of the IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom). 2016 Presented at: IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom); 8-10 Oct. 2016; Atlanta, GA, USA p. 201-208. [doi: 10.1109/bdcloud-socialcom-sustaincom.2016.40]
161. Martin CK, Nicklas T, Gunturk B, Correa JB, Allen HR, Champagne C. Measuring food intake with digital photography. J Hum Nutr Diet 2014 Jan;27 Suppl 1:72-81 [FREE Full text] [doi: 10.1111/jhn.12014] [Medline: 23848588]
162. Martin WH, Griest SE, Sobel JL, Howarth LC. Randomized trial of four noise-induced hearing loss and tinnitus prevention interventions for children. Int J Audiol 2013 Feb 04;52 Suppl 1(sup1):S41-S49. [doi: 10.3109/14992027.2012.743048] [Medline: 23373742]
163. McCormack GR, McFadden K, McHugh TF, Spence JC, Mummery K. Barriers and facilitators impacting the experiences of adults participating in an internet-facilitated pedometer intervention. Psychology of Sport and Exercise 2019 Nov;45:101549. [doi: 10.1016/j.psychsport.2019.101549]
164. McGloin AF, Eslami S. Digital and social media opportunities for dietary behaviour change. Proc Nutr Soc 2015 May;74(2):139-148. [doi: 10.1017/S0029665114001505] [Medline: 25319345]
165. Meng Y, Kim HC. Towards design of technologies persuading more physical activity. In: Proceedings of the IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013). 2013 Presented at: IEEE 15th International Conference on e-Health Networking, Applications and Services (Healthcom 2013); 9-12 Oct. 2013; Lisbon. [doi: 10.1109/healthcom.2013.6720768]
166. Mitchell BL, Lewis NR, Smith AE, Rowlands AV, Parfitt G, Dollman J. Rural Environments and Community Health (REACH): a randomised controlled trial protocol for an online walking intervention in rural adults. BMC Public Health 2014 Sep 18;14(1):969 [FREE Full text] [doi: 10.1186/1471-2458-14-969] [Medline: 25236776]
167. Mitchell BL, Smith AE, Rowlands AV, Fraysse F, Parfitt G, Lewis NR, et al. Promoting physical activity in rural Australian adults using an online intervention. J Sci Med Sport 2019 Jan;22(1):70-75. [doi: 10.1016/j.jsams.2018.07.002] [Medline: 30031747]
168. Money AG, Atwal A, Boyce E, Gaber S, Windeatt S, Alexandrou K. Falls Sensei: a serious 3D exploration game to enable the detection of extrinsic home fall hazards for older adults. BMC Med Inform Decis Mak 2019 Apr 16;19(1):85 [FREE Full text] [doi: 10.1186/s12911-019-0808-x] [Medline: 30992072]
169. Nahm E, Resnick B, Bellantoni M, Zhu S, Brown C, Brennan PF, et al. Dissemination of a theory-based online bone health program: Two intervention approaches. Health Informatics J 2015 Jun 28;21(2):120-136 [FREE Full text] [doi: 10.1177/1460458213505573] [Medline: 26021668]
170. Neiger BL, Thackeray R, Burton SH, Giraud-Carrier CG, Fagen MC. Evaluating social media's capacity to develop engaged audiences in health promotion settings: use of Twitter metrics as a case study. Health Promot Pract 2013 Mar;14(2):157-162. [doi: 10.1177/1524839912469378] [Medline: 23271716]
171. Newton NC, Teesson M, Vogl LE, Andrews G. Internet-based prevention for alcohol and cannabis use: final results of the Climate Schools course. Addiction 2010 Apr;105(4):749-759. [doi: 10.1111/j.1360-0443.2009.02853.x] [Medline: 20148791]

172. Nitsch M, Adamcik T, Kuso S, Zeiler M, Waldherr K. Usability and Engagement Evaluation of an Unguided Online Program for Promoting a Healthy Lifestyle and Reducing the Risk for Eating Disorders and Obesity in the School Setting. Nutrients 2019 Mar 27;11(4):713 [FREE Full text] [doi: 10.3390/nu11040713] [Medline: 30934754]
173. O'Mara B. Social media, digital video and health promotion in a culturally and linguistically diverse Australia. Health Promot Int 2013 Sep 30;28(3):466-476. [doi: 10.1093/heapro/das014] [Medline: 22466682]
174. Paschall MJ, Antin T, Ringwalt CL, Saltz RF. Evaluation of an Internet-based alcohol misuse prevention course for college freshmen: findings of a randomized multi-campus trial. Am J Prev Med 2011 Sep;41(3):300-308 [FREE Full text] [doi: 10.1016/j.amepre.2011.03.021] [Medline: 21855745]
175. Payton F, Kvasny L. Online HIV awareness and technology affordance benefits for black female collegians - maybe not: the case of stigma. J Am Med Inform Assoc 2016 Nov;23(6):1121-1126. [doi: 10.1093/jamia/ocw017] [Medline: 27094988]
176. Peels DA, de Vries H, Bolman C, Golsteijn RHJ, van Stralen MM, Mudde AN, et al. Differences in the use and appreciation of a web-based or printed computer-tailored physical activity intervention for people aged over 50 years. Health Educ Res 2013 Aug 18;28(4):715-731. [doi: 10.1093/her/cyt065] [Medline: 23784076]
177. Peels DA, van Stralen MM, Bolman C, Golsteijn RH, de Vries H, Mudde AN, et al. Development of web-based computer-tailored advice to promote physical activity among people older than 50 years. J Med Internet Res 2012 Mar 02;14(2):e39 [FREE Full text] [doi: 10.2196/jmir.1742] [Medline: 22390878]
178. Peskin MF, Shegog R, Markham CM, Thiel M, Baumler ER, Addy RC, et al. Efficacy of It's Your Game-Tech: A Computer-Based Sexual Health Education Program for Middle School Youth. J Adolesc Health 2015 May;56(5):515-521 [FREE Full text] [doi: 10.1016/j.jadohealth.2015.01.001] [Medline: 25739520]
179. Phan N, Ebrahimi J, Kil D, Piniewski B, Dou D. Topic-Aware Physical Activity Propagation in a Health Social Network. IEEE Intell. Syst 2016 Jan;31(1):5-14. [doi: 10.1109/mis.2015.92]
180. Plaete J, Crombez G, Van der Mispel C, Verloigne M, Van Stappen V, De Bourdeaudhuij I. Effect of the Web-Based Intervention MyPlan 1.0 on Self-Reported Fruit and Vegetable Intake in Adults Who Visit General Practice: A Quasi-Experimental Trial. J Med Internet Res 2016 Feb 29;18(2):e47 [FREE Full text] [doi: 10.2196/jmir.5252] [Medline: 26929095]
181. Poppe L, Plaete J, Huys N, Verloigne M, Deveugele M, De Bourdeaudhuij I, et al. Process Evaluation of an eHealth Intervention Implemented into General Practice: General Practitioners' and Patients' Views. Int J Environ Res Public Health 2018 Jul 12;15(7):1475 [FREE Full text] [doi: 10.3390/ijerph15071475] [Medline: 30002338]
182. Prins RG, van Empelen P, Beenackers MA, Brug J, Oenema A. Systematic Development of the YouRAction program, a computer-tailored physical activity promotion intervention for Dutch adolescents, targeting personal motivations and environmental opportunities. BMC Public Health 2010 Aug 11;10(1):474 [FREE Full text] [doi: 10.1186/1471-2458-10-474] [Medline: 20701782]
183. Prins RG, Brug J, van Empelen P, Oenema A. Effectiveness of YouRAction, an intervention to promote adolescent physical activity using personal and environmental feedback: a cluster RCT. PLoS One 2012 Mar 5;7(3):e32682 [FREE Full text] [doi: 10.1371/journal.pone.0032682] [Medline: 22403695]
184. Raghupathy S, Forth ALG. The HAWK2 program: a computer-based drug prevention intervention for Native American youth. Am J Drug Alcohol Abuse 2012 Sep;38(5):461-467. [doi: 10.3109/00952990.2012.694531] [Medline: 22823599]
185. Reed-Jones RJ, Dorgo S, Hitchings MK, Bader JO. WiiFit™ Plus balance test scores for the assessment of balance and mobility in older adults. Gait Posture 2012 Jul;36(3):430-433 [FREE Full text] [doi: 10.1016/j.gaitpost.2012.03.027] [Medline: 22534562]
186. Rice E. The positive role of social networks and social networking technology in the condom-using behaviors of homeless young people. Public Health Rep 2010 Jul 01;125(4):588-595 [FREE Full text] [doi: 10.1177/003335491012500414] [Medline: 20597459]
187. Rice E, Milburn NG, Monro W. Social networking technology, social network composition, and reductions in substance use among homeless adolescents. Prev Sci 2011 Mar 31;12(1):80-88 [FREE Full text] [doi: 10.1007/s11121-010-0191-4] [Medline: 21194011]
188. Rice E, Tulbert E, Cederbaum J, Barman Adhikari A, Milburn NG. Mobilizing homeless youth for HIV prevention: a social network analysis of the acceptability of a face-to-face and online social networking intervention. Health Educ Res 2012 Apr 13;27(2):226-236 [FREE Full text] [doi: 10.1093/her/cyr113] [Medline: 22247453]
189. Rith-Najarian L, Sun W, Chen A, Chorpita B, Chavira D, Mougalian S, et al. What's in a name? Branding of online mental health programming for university students. J Consult Clin Psychol 2019 Apr;87(4):380-391. [doi: 10.1037/ccp0000383] [Medline: 30883165]
190. Robertson J, Jepson R, Macvean A, Gray S. Understanding the Importance of Context: A Qualitative Study of a Location-Based Exergame to Enhance School Childrens Physical Activity. PLoS One 2016 Aug 22;11(8):e0160927 [FREE Full text] [doi: 10.1371/journal.pone.0160927] [Medline: 27548504]
191. Rote AE, Klos LA, Brondino MJ, Harley AE, Swartz AM. The Efficacy of a Walking Intervention Using Social Media to Increase Physical Activity: A Randomized Trial. J Phys Act Health 2015 Jun;12 Suppl 1:S18-S25. [doi: 10.1123/jpah.2014-0279] [Medline: 25599378]

192. Rovniak LS, Sallis JF, Kraschnewski JL, Sciamanna CN, Kiser EJ, Ray CA, et al. Engineering online and in-person social networks to sustain physical activity: application of a conceptual model. BMC Public Health 2013 Aug 14;13(1):753 [FREE Full text] [doi: 10.1186/1471-2458-13-753] [Medline: 23945138]
193. Schoene D, Lord SR, Delbaere K, Severino C, Davies TA, Smith ST. A randomized controlled pilot study of home-based step training in older people using videogame technology. PLoS One 2013 Mar 5;8(3):e57734 [FREE Full text] [doi: 10.1371/journal.pone.0057734] [Medline: 23472104]
194. Schwinn TM, Schinke S, Fang L, Kandasamy S. A web-based, health promotion program for adolescent girls and their mothers who reside in public housing. Addict Behav 2014 Apr;39(4):757-760 [FREE Full text] [doi: 10.1016/j.addbeh.2013.11.029] [Medline: 24447886]
195. Shafii T, Benson SK, Morrison DM, Hughes JP, Golden MR, Holmes KK. Results from e-KISS: electronic-KIOSK Intervention for Safer Sex: A pilot randomized controlled trial of an interactive computer-based intervention for sexual health in adolescents and young adults. PLoS One 2019 Jan 23;14(1):e0209064 [FREE Full text] [doi: 10.1371/journal.pone.0209064] [Medline: 30673710]
196. Shanta Bridges L, Sharma M, Lee JHSH, Bennett R, Buxbaum SG, Reese-Smith J. Using the PRECEDE-PROCEED model for an online peer-to-peer suicide prevention and awareness for depression (SPAD) intervention among African American college students: experimental study. Health Promot Perspect 2018 Dec 30;8(1):15-24 [FREE Full text] [doi: 10.15171/hpp.2018.02] [Medline: 29423358]
197. Sharma SV, Shegog R, Chow J, Finley C, Pomeroy M, Smith C, et al. Effects of the Quest to Lava Mountain Computer Game on Dietary and Physical Activity Behaviors of Elementary School Children: A Pilot Group-Randomized Controlled Trial. J Acad Nutr Diet 2015 Aug;115(8):1260-1271. [doi: 10.1016/j.jand.2015.02.022] [Medline: 25912520]
198. Shegog R, Craig RS, Gorman G, Jessen C, Torres J, Lane TL, et al. NATIVE-It's Your Game: Adapting a Technology-Based Sexual Health Curriculum for American Indian and Alaska Native youth. J Prim Prev 2017;38(1-2):27-48. [doi: 10.1007/s10935-016-0440-9] [Medline: 27520459]
199. Smith M, Denali DL. Social Media in Health Education, Promotion, and Communication: Reaching Rural Hispanic Populations along the USA/Mexico Border Region. J. Racial and Ethnic Health Disparities 2014 Jun 18;1(3):194-198. [doi: 10.1007/s40615-014-0025-3]
200. West DS, Monroe CM, Turner-McGrievy G, Sundstrom B, Larsen C, Magradey K, et al. A Technology-Mediated Behavioral Weight Gain Prevention Intervention for College Students: Controlled, Quasi-Experimental Study. J Med Internet Res 2016 Jun 13;18(6):e133 [FREE Full text] [doi: 10.2196/jmir.5474] [Medline: 27296086]
201. Springvloet L, Lechner L, Oenema A. Planned development and evaluation protocol of two versions of a web-based computer-tailored nutrition education intervention aimed at adults, including cognitive and environmental feedback. BMC Public Health 2014 Jan 17;14:47 [FREE Full text] [doi: 10.1186/1471-2458-14-47] [Medline: 24438381]
202. Suomi R, Mäntymäki M, Söderlund S. Promoting health in virtual worlds: lessons from second life. J Med Internet Res 2014 Oct 13;16(10):e229 [FREE Full text] [doi: 10.2196/jmir.3177] [Medline: 25313009]
203. Tague R, Maeder A, Vandelanotte C, Kolt G, Caperchione C, Rosenkranz R, et al. Assessing user engagement in a health promotion website using social networking. Studies in Health Technology and Informatics 2014;206:84-92. [doi: 10.3233/978-1-61499-456-5-84]
204. Digital Active Break. Techniker Krankenkasse. 2019. URL: https://www.tk.de/presse/themen/praevention/gesund-leben/dab-projektfsu-jena-2068254 [accessed 2021-12-23]
205. Teitelman AM, Kim SK, Waas R, DeSenna A, Duncan R. Development of the NowIKnow Mobile Application to Promote Completion of HPV Vaccine Series Among Young Adult Women. J Obstet Gynecol Neonatal Nurs 2018 Nov;47(6):844-852. [doi: 10.1016/j.jogn.2018.06.001] [Medline: 30036509]
206. Tolks D, Sailer M, Dadaczynski K, Lampert C, Huberty J, Paulus P, et al. ONYA—The Wellbeing Game: How to Use Gamification to Promote Wellbeing. Information 2019 Feb 12;10(2):58. [doi: 10.3390/info10020058]
207. Tortolero SR, Markham CM, Peskin MF, Shegog R, Addy RC, Escobar-Chaves SL, et al. It's Your Game: Keep It Real: delaying sexual behavior with an effective middle school program. J Adolesc Health 2010 Feb;46(2):169-179 [FREE Full text] [doi: 10.1016/j.jadohealth.2009.06.008] [Medline: 20113923]
208. Tsai T, Wong AM, Hsu C, Tseng KC. Research on a community-based platform for promoting health and physical fitness in the elderly community. PLoS One 2013 Feb 27;8(2):e57452 [FREE Full text] [doi: 10.1371/journal.pone.0057452] [Medline: 23460859]
209. Van Kessel G, Kavanagh M, Maher C. A Qualitative Study to Examine Feasibility and Design of an Online Social Networking Intervention to Increase Physical Activity in Teenage Girls. PLoS One 2016 Mar 2;11(3):e0150817 [FREE Full text] [doi: 10.1371/journal.pone.0150817] [Medline: 26934191]
210. Van Lippevelde W, Vangeel J, De Cock N, Lachat C, Goossens L, Beullens K, et al. Using a gamified monitoring app to change adolescents' snack intake: the development of the REWARD app and evaluation design. BMC Public Health 2016 Aug 05;16(1):725 [FREE Full text] [doi: 10.1186/s12889-016-3286-4] [Medline: 27494932]
211. van Stralen MM, de Vries H, Bolman C, Mudde AN, Lechner L. Exploring the efficacy and moderators of two computer-tailored physical activity interventions for older adults: a randomized controlled trial. Ann Behav Med 2010 May 25;39(2):139-150. [doi: 10.1007/s12160-010-9166-8] [Medline: 20182833]

212. van Stralen MM, de Vries H, Mudde AN, Bolman C, Lechner L. The long-term efficacy of two computer-tailored physical activity interventions for older adults: main effects and mediators. Health Psychol 2011 Jul;30(4):442-452. [doi: 10.1037/a0023579] [Medline: 21639638]
213. Walton MA, Resko S, Barry KL, Chermack ST, Zucker RA, Zimmerman MA, et al. A randomized controlled trial testing the efficacy of a brief cannabis universal prevention program among adolescents in primary care. Addiction 2014 May;109(5):786-797 [FREE Full text] [doi: 10.1111/add.12469] [Medline: 24372937]
214. Ward DS, Vaughn AE, Mazzucca S, Burney R. Translating a child care based intervention for online delivery: development and randomized pilot study of Go NAPSACC. BMC Public Health 2017 Nov 21;17(1):891 [FREE Full text] [doi: 10.1186/s12889-017-4898-z] [Medline: 29162057]
215. Washington E, Weed L, Vardaman S. Using The Internet To Increase Physical Activity In A Faith Community. Journal of Christian Nursing 2015;32(3):168-173. [doi: 10.1097/cnj.0000000000000180]
216. Watkins DC. Improving the Living, Learning, and Thriving of Young Black Men: A Conceptual Framework for Reflection and Projection. Int J Environ Res Public Health 2019 Apr 13;16(8):1331 [FREE Full text] [doi: 10.3390/ijerph16081331] [Medline: 31013882]
217. Watkins DC, Allen JO, Goodwill JR, Noel B. Strengths and weaknesses of the Young Black Men, Masculinities, and Mental Health (YBMen) Facebook project. Am J Orthopsychiatry 2017;87(4):392-401. [doi: 10.1037/ort0000229] [Medline: 27977287]
218. Webster R, Gerressu M, Michie S, Estcourt C, Anderson J, Ang CS, et al. Defining the Content of an Online Sexual Health Intervention: The MenSS Website. JMIR Res Protoc 2015 Jul;4(3):e82 [FREE Full text] [doi: 10.2196/resprot.4316] [Medline: 26142304]
219. Whittemore R, Chao A, Jang M, Jeon S, Liptak T, Popick R, et al. Implementation of a school-based internet obesity prevention program for adolescents. J Nutr Educ Behav 2013 Nov;45(6):586-594 [FREE Full text] [doi: 10.1016/j.jneb.2013.03.012] [Medline: 23850014]
220. Widman L, Golin C, Kamke K, Massey J, Prinstein M. Feasibility and acceptability of a web-based HIV/STD prevention program for adolescent girls targeting sexual communication skills. Health Educ Res 2017 Aug 01;32(4):343-352 [FREE Full text] [doi: 10.1093/her/cyx048] [Medline: 28854574]
221. Widman L, Golin CE, Kamke K, Burnette JL, Prinstein MJ. Sexual Assertiveness Skills and Sexual Decision-Making in Adolescent Girls: Randomized Controlled Trial of an Online Program. Am J Public Health 2018 Jan;108(1):96-102. [doi: 10.2105/AJPH.2017.304106] [Medline: 29161072]
222. Widman L, Kamke K, Evans R, Stewart JL, Choukas-Bradley S, Golin CE. Feasibility, Acceptability, and Preliminary Efficacy of a Brief Online Sexual Health Program for Adolescents. J Sex Res 2020 Feb 09;57(2):145-154 [FREE Full text] [doi: 10.1080/00224499.2019.1630800] [Medline: 31287336]
223. Winskell K, Sabben G, Ondeng'e K, Odero I, Akelo V, Mudhune V. A Smartphone Game to Prevent HIV among Young Kenyans: Household Dynamics of Gameplay in a Feasibility Study. Health Educ J 2019 Aug 01;78(5):595-606 [FREE Full text] [doi: 10.1177/0017896919832344] [Medline: 34219796]
224. Witzel TC, Guise A, Nutland W, Bourne A. It Starts With Me: Privacy concerns and stigma in the evaluation of a Facebook health promotion intervention. Sex Health 2016 Jun;13(3):228-233. [doi: 10.1071/SH15231] [Medline: 27075940]
225. Wright CJC, Schwarzman J, Dietze PM, Crockett B, Lim MSC. Barriers and opportunities in the translation of mobile phone and social media interventions between research and health promotion practice in Australia: a qualitative study of expert perspectives. Health Res Policy Syst 2019 Jan 10;17(1):5 [FREE Full text] [doi: 10.1186/s12961-018-0406-x] [Medline: 30630497]
226. Ybarra ML, Biringi R, Prescott T, Bull SS. Usability and Navigability of an HIV/AIDS Internet Intervention for Adolescents in a Resource-Limited Setting. CIN: Computers, Informatics, Nursing 2012;30(11):587-595. [doi: 10.1097/nxn.0b013e318266cb0e]
227. Yepes MF. Mobile Tablet Menus. Cornell Hospitality Quarterly 2014 Aug 28;56(1):58-67. [doi: 10.1177/1938965514546371]
228. Yoost JL, Starcher RW, King-Mallory RA, Hussain N, Hensley CA, Gress TW. The Use of Telehealth to Teach Reproductive Health to Female Rural High School Students. J Pediatr Adolesc Gynecol 2017 Apr;30(2):193-198. [doi: 10.1016/j.jpag.2016.10.002] [Medline: 27742427]
229. Youm S, Park S. How the awareness of u-healthcare service and health conditions affect healthy lifestyle: an empirical analysis based on a u-healthcare service experience. Telemed J E Health 2015 Apr;21(4):286-295 [FREE Full text] [doi: 10.1089/tmj.2014.0092] [Medline: 25635473]
230. Young SD, Cumberland WG, Lee S, Jaganath D, Szekeres G, Coates T. Social networking technologies as an emerging tool for HIV prevention: a cluster randomized trial. Ann Intern Med 2013 Sep 3;159(5):318-324 [FREE Full text] [doi: 10.7326/0003-4819-159-5-201309030-00005] [Medline: 24026317]
231. Young SD, Harrell L, Jaganath D, Cohen AC, Shoptaw S. Feasibility of recruiting peer educators for an online social networking-based health intervention. Health Educ J 2013 May;72(3):276-282 [FREE Full text] [doi: 10.1177/0017896912440768] [Medline: 25530624]

232. Young SD, Holloway I, Jaganath D, Rice E, Westmoreland D, Coates T. Project HOPE: Online Social Network Changes in an HIV Prevention Randomized Controlled Trial for African American and Latino Men Who Have Sex With Men. Am J Public Health 2014 Sep;104(9):1707-1712. [doi: 10.2105/ajph.2014.301992]
233. Zhang J, Brackbill D, Yang S, Centola D. Efficacy and causal mechanism of an online social media intervention to increase physical activity: Results of a randomized controlled trial. Prev Med Rep 2015;2:651-657 [FREE Full text] [doi: 10.1016/j.pmedr.2015.08.005] [Medline: 26844132]
234. Zhang J, Brackbill D, Yang S, Becker J, Herbert N, Centola D. Support or competition? How online social networks increase physical activity: A randomized controlled trial. Prev Med Rep 2016 Dec;4:453-458 [FREE Full text] [doi: 10.1016/j.pmedr.2016.08.008] [Medline: 27617191]
235. Zhao W, Lun R. A Kinect-based system for promoting healthier living at home. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics (SMC). 2016 Presented at: IEEE International Conference on Systems, Man, and Cybernetics (SMC); 9-12 Oct. 2016; Budapest, Hungary p. 258-263. [doi: 10.1109/smc.2016.7844251]
236. Norman GJ, Zabinski MF, Adams MA, Rosenberg DE, Yaroch AL, Atienza AA. A review of eHealth interventions for physical activity and dietary behavior change. Am J Prev Med 2007 Oct;33(4):336-345 [FREE Full text] [doi: 10.1016/j.amepre.2007.05.007] [Medline: 17888860]
237. Bailey J, Mann S, Wayal S, Hunter R, Free C, Abraham C, et al. Sexual health promotion for young people delivered via digital media: a scoping review. Public Health Res 2015 Nov;3(13):1-120. [doi: 10.3310/phr03130] [Medline: 26583166]
238. Lupton D. Health promotion in the digital era: a critical commentary. Health Promot Int 2015 Mar 15;30(1):174-183. [doi: 10.1093/heapro/dau091] [Medline: 25320120]
239. Taggart T, Grewe ME, Conserve DF, Gliwa C, Roman IM. Social Media and HIV: A Systematic Review of Uses of Social Media in HIV Communication. J Med Internet Res 2015 Nov;17(11):e248 [FREE Full text] [doi: 10.2196/jmir.4387] [Medline: 26525289]
240. Stellefson M, Paige SR, Chaney BH, Chaney JD. Evolving Role of Social Media in Health Promotion: Updated Responsibilities for Health Education Specialists. Int J Environ Res Public Health 2020 Feb 12;17(4):1153 [FREE Full text] [doi: 10.3390/ijerph17041153] [Medline: 32059561]
241. Vedel I, Ramaprasad J, Lapointe L. Social Media Strategies for Health Promotion by Nonprofit Organizations: Multiple Case Study Design. J Med Internet Res 2020 Apr 06;22(4):e15586 [FREE Full text] [doi: 10.2196/15586] [Medline: 32250282]
242. Nour M, Chen J, Allman-Farinelli M. Efficacy and External Validity of Electronic and Mobile Phone-Based Interventions Promoting Vegetable Intake in Young Adults: Systematic Review and Meta-Analysis. J Med Internet Res 2016 Apr 08;18(4):e58 [FREE Full text] [doi: 10.2196/jmir.5082] [Medline: 27059765]
243. Healthy settings. World Health Organization. URL: https://www.who.int/teams/health-promotion/enhanced-wellbeing/healthy-settings [accessed 2021-12-01]

Abbreviations

mHealth: mobile health
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews
RCT: randomized controlled trial
WHO: World Health Organization

Edited by A Mavragani; submitted 06.10.20; peer-reviewed by A Hidki, S Tomczyk; comments to author 03.12.20; revised version received 16.12.20; accepted 02.12.21; published 28.01.22.

Please cite as:
Stark AL, Geukes C, Dockweiler C
Digital Health Promotion and Prevention in Settings: Scoping Review
J Med Internet Res 2022;24(1):e21063
URL: https://www.jmir.org/2022/1/e21063
doi:10.2196/21063
PMID:35089140

Viewpoint

Designing Effective eHealth Interventions for Underserved Groups: Five Lessons From a Decade of eHealth Intervention Design and Deployment

Edmund WJ Lee[1], BA, MA, PhD; Rachel F McCloud[2], MPH, ScD; Kasisomayajula Viswanath[2,3], PhD

[1]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

[2]Dana-Farber Cancer Institute, Boston, MA, United States

[3]Harvard T H Chan School of Public Health, Boston, MA, United States

Corresponding Author:
Edmund WJ Lee, BA, MA, PhD
Wee Kim Wee School of Communication and Information
Nanyang Technological University
31 Nanyang Link
Wee Kim Wee School of Communication and Information Building
Singapore, 637718
Singapore
Phone: 65 69083429
Email: elwj88@gmail.com

Abstract

Despite the proliferation of eHealth interventions, such as web portals, for health information dissemination or the use of mobile apps and wearables for health monitoring, research has shown that underserved groups do not benefit proportionately from these eHealth interventions. This is largely because of usability issues and the lack of attention to the broader structural, physical, and psychosocial barriers to technology adoption and use. The objective of this paper is to draw lessons from a decade of experience in designing different user-centered eHealth interventions (eg, web portals and health apps) to inform future work in leveraging technology to address health disparities. We draw these lessons from a series of interventions from the work we have done over 15 years in the Viswanath laboratory at the Dana-Farber Cancer Institute and Harvard TH Chan School of Public Health, focusing on three projects that used web portals and health apps targeted toward underserved groups. The projects were the following: *Click to Connect*, which was a community-based eHealth intervention that aimed to improve internet skills and health literacy among underserved groups by providing home access to high-speed internet, computer, and internet training classes, as well as a dedicated health web portal with ongoing technical support; *PLANET MassCONECT*, which was a knowledge translation project that built capacity among community-based organizations in Boston, Lawrence, and Worcester in Massachusetts to adopt evidence-based health promotion programs; and *Smartphone App for Public Health*, which was a mobile health research that facilitated both participatory (eg, surveys) and passive data (eg, geolocations and web-browsing behaviors) collection for the purpose of understanding tobacco message exposure in individuals' built environment. Through our work, we distilled five key principles for researchers aiming to design eHealth interventions for underserved groups. They are as follows: develop a strategic road map to address communication inequalities (ie, a concrete action plan to identify the barriers faced by underserved groups and customize specific solutions to each of them), engage multiple stakeholders from the beginning for the long haul, design with usability—readability and navigability—in mind, build privacy safeguards into eHealth interventions and communicate privacy–utility tradeoffs in simplicity, and strive for an optimal balance between open science aspirations and protection of underserved groups.

(J Med Internet Res 2022;24(1):e25419) doi:10.2196/25419

KEYWORDS

eHealth; mobile health; communication inequalities; health disparities; health informatics; mobile phone

Background

Breakthroughs in communication technologies in the past decade have brought about a plethora of opportunities for researchers to leverage different forms of digital technologies to bridge health disparities. With the ubiquitous use of digital tools such as computers, laptops, and smartphones and recent advances in cyberinfrastructure to store and analyze big data at scale with efficiency, researchers have piloted and designed web portals to specifically aid health-information seeking among different populations [1,2], encouraged the use of patient portals for health care management [3], and developed smartphone apps and wearables to improve health monitoring [4].

eHealth Interventions in Widening Inequality

Despite the benefits brought about by the communications revolution, a major critique in eHealth intervention research is that instead of closing the gaps between the rich and the poor, the introduction of digital gadgets reproduces and reinforces existing social structures of health inequalities [5-8]. For instance, research has shown that in the United States, compared with White and non-Hispanic individuals, Black and Hispanic individuals were significantly less likely to be offered access to web-based personal health information portals by their health care providers, and they were also less likely to use personal health information, even when granted access [9]. Although the combination of artificial intelligence algorithms and the availability of electronic health records (EHRs) promises to enable clinicians to predict hospital readmissions and mortality [10], people from lower socioeconomic groups may not necessarily benefit from these if their health records are not recorded in the EHRs in the first place because of multiple missed appointments [11]. In some instances, the places where people from lower socioeconomic groups seek treatment may not be equipped with EHRs. Even when people from lower socioeconomic groups have access to digital health technologies, they may not be able to afford the recurring expenses needed to use the services or maintain stable internet connections [12]. Even if access to these technologies is not an issue, underserved groups may find it difficult to navigate the complex platforms of different digital health technologies [13].

Beyond financial and technology factors, research has shown that another significant challenge to eHealth interventions is the lack of collective use and adoption by people's social networks [6]. A clear example is in the current COVID-19 pandemic, where a major challenge in implementing and leveraging data from contact tracing apps is the low uptake by the general population [14]. The problem is compounded when people from marginalized communities and people of color, who have been hit the hardest by the pandemic, remain skeptical of technology because of the historical baggage of being unfairly targeted by state surveillance or being wary of the amount of misinformation and disinformation generated from these digital sources [15,16].

In today's world, where data collected from eHealth interventions can be used to understand health behaviors when individuals from underserved groups do not use the digital gadgets designed for them and are systematically left out, it contributes to a larger problem called *data absenteeism*—missing data of underserved groups in health systems—leading to inaccuracies and perpetuation of biases when training machine learning models on these data sets with large amounts of missing data [17].

The Need to Address Inequality in eHealth Interventions

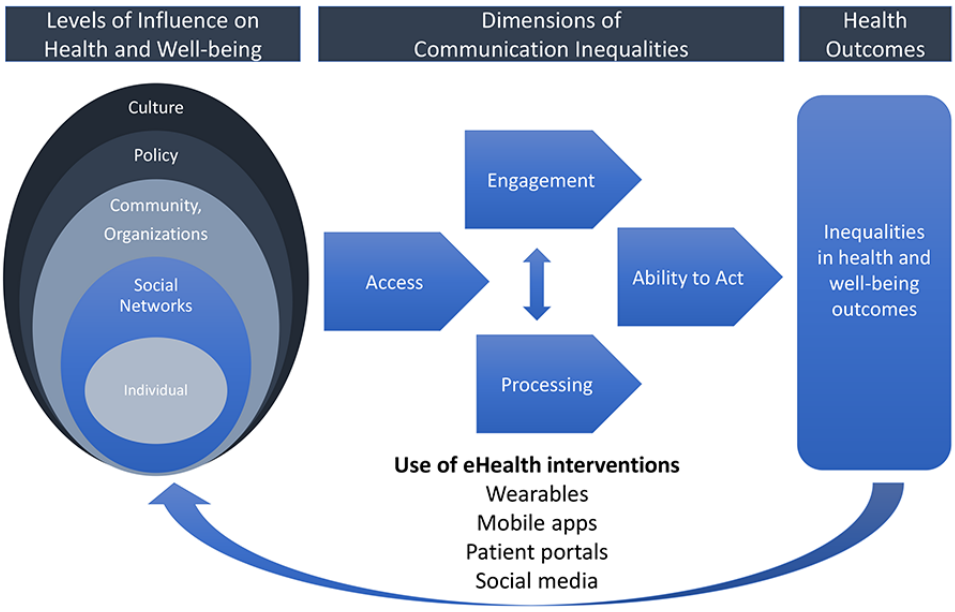
To level the playing field in ensuring that eHealth interventions benefit underserved groups, it is paramount that researchers and public health organizations pay attention to the underlying mechanisms that contribute to *communication inequalities*. Communication inequalities are structural, interpersonal, and individual differences among social groups in accessing, using, and processing information from media and communication technologies, and they mediate the relationship between social determinants and health outcomes [18]. Thus, although investing in the usability of eHealth interventions is important, it is crucial that public health scholars go beyond technological considerations and examine how external, social, and individual contexts synergistically and collectively influence the success of eHealth interventions [17]. After all, decades of public health research have fervently called for the need to engage individuals, social networks, the broader community, and anchor institutions (eg, hospitals and community health care organizations) in implementing health promotion programs [19,20]. The purpose of this paper is to showcase and draw lessons from over a decade of our research experience in implementing different eHealth interventions among underserved groups.

A Framework for Addressing Communication Inequalities: Structural Influence Model

Our work has focused on examining information and technologies that can help bridge unequal health outcomes between different social groups, guided primarily by the structural influence model of communication (Figure 1), which delineates the relationship between how communication inequalities mediate and moderate the association between social determinants and communication and health outcomes [18,21]. The framework postulates that the use of eHealth interventions can have both direct and indirect impacts on health, and they may potentially amplify disparities among different social, racial, or ethnic groups. Central to the framework is the concept of communication inequalities, which is defined as unequal access and use of communication technologies because of differences in how people access, use, and process information from digital media platforms, which widens existing health disparities [7,17,22]. In other words, the fundamental assumption is that whoever has the power as well as the cognitive, social, and financial resources (typically, people with higher income and education) to access, use, and process information from

communication technologies is most likely to benefit, whereas groups that are underserved, even if they have access to such platforms, seldom reap the promised benefits though they are often the target audience of such interventions [23].

Figure 1. Modified structural influence model.



Despite the stark reality of health disparities, the silver lining is that communication inequalities are ultimately *modifiable* conditions that could be addressed by academic institutions, public health organizations, or government agencies more readily as compared with social determinants. To be effective, the design of eHealth interventions should consider broader contexts that may attenuate the uptake and use of communication technologies.

At the individual level, research has shown that underserved groups face significantly more barriers— economically, socially, and technologically—in the adoption of eHealth interventions when compared with their financially better off counterparts [24]. A study on the adoption of web-based patient portals found that although participants were given the opportunity to register for web-based patient portals, only 22% of patients with limited health literacy registered for the account. This was lower when compared with 73% of adults with adequate health literacy skills [25]. This pattern was consistent with racial differences as well, with minority groups being at higher odds of not using these eHealth interventions [26]. Beyond individual factors, external factors, such as individuals' social networks, the type of hospitals in the community, and community-based organizations (CBOs) play a crucial role in boosting health literacy [27-32]. Thus, it will be strategic to gain insights into CBOs and social networks of the underserved groups, as they would have the experience in identifying the type of interventions that would best suit the needs of the underserved populations and those that would simply not work [33,34]. Without the input of the community, designers of eHealth interventions may fall into the trap of having *solutions in search of a problem* [35] and forcing innovations that simply do not work in the community.

The rest of this paper will illustrate the key lessons learned from three signature projects in the Viswanath laboratory. The first project was called Click to Connect (C2C), which was a community-based eHealth intervention that aimed to improve internet skills and health literacy among underserved groups by providing home access to high-speed internet, computer, and internet training classes, as well as a dedicated health web portal with ongoing technical support. The second project was PLANET MassCONECT, which was a knowledge translation project that built capacity among CBOs in Boston, Lawrence, and Worcester in Massachusetts to adopt evidence-based health promotion programs. Finally, the third project was Project Smartphone App for Public Health (SNAP), which was a mobile health research that collected both participatory (eg, surveys) and passive data (eg, geolocations and web-browsing behaviors) for the purpose of understanding tobacco message exposure in an individual's built environment.

We have distilled five key principles on eHealth intervention research for underserved groups from these three projects: (1) develop a strategic road map to address communication inequalities (ie, a concrete action plan to identify the barriers faced by underserved groups and customizing specific solutions to each of them), (2) engage multiple stakeholders from the beginning for the long haul, (3) design with usability—readability and navigability—in mind, (4) build privacy safeguards into eHealth interventions and communicate privacy–utility tradeoffs in simplicity, and (5) strive for an optimal balance between open science aspirations and protection of underserved groups. We have also proposed multilevel strategies for developing eHealth interventions that health

organizations could draw on in their work with underserved groups.

Project 1: C2C

Project C2C was a randomized controlled trial that aimed to improve eHealth literacy among people from lower socioeconomic positions (SEPs). To achieve the objective of empowering people from lower SEP groups by taking advantage of web-based health portals to seek information and gain health knowledge, we designed an intervention that involved the (1) development of a web-based health portal (Figure 2) from scratch that was customized for novice or less experienced users to easily navigate and access the internet, specifically health information; (2) purchase and provision of computer and broadband internet access for the entire length of study; (3) training classes where participants were taught digital skills such as how to use computers and the internet; and (4) ongoing technical support if participants had any questions on the health web portal or connectivity issues [34]. The trial was conducted in three waves from 2007 to 2009, and for each wave, participants attended 9 monthly training classes at community colleges located in Boston. Participants randomized to the control group received health information at the end of the study period. Participants were recruited from adult education centers located in the Greater Boston area of the state of Massachusetts [36].

Figure 2. Screenshot of the web-based health portal of Click to Connect.

Researchers generally make insufficient efforts to reach underserved populations with the stereotype that they are *hard to reach*. We adopted a *proactive* approach, which involved arranging face-to-face contact with community leaders and organizations, as well as recruitment presentations and meetings in the community [37]. Overall, we made a total of 190 in-person presentations at 32 adult literacy centers in Greater Boston from May 2007 to October 2009.

Project 2: PLANET MassCONECT

Although project C2C was targeted at individuals from underserved groups, project PLANET MassCONECT was a 5-year knowledge translation project funded by the US National Cancer Institute that aimed to build the capacity to find, adapt, and implement evidence-based health promotion programs among a diverse group of CBOs located in three cities in the state of Massachusetts (Boston, Lawrence, and Worcester) in the United States [38]. In other words, PLANET MassCONECT aimed to strengthen CBOs by creating a network where organizations could collaborate with other CBOs, researchers, and community members and provide training and technical support to use a customized web-based health portal to find relevant health resources and data to implement evidence-based health promotion projects. At the end of the training, CBOs were plugged into a network with other CBOs and encouraged to apply for mini-grants to strengthen interorganizational collaboration in addressing health disparities [39].

These aims were achieved through an intervention that included these main components: (1) development of a web-based health portal containing health resources, including evidence-based interventions, data, and resources for interventions; (2) a 2-day capacity-building workshop for CBOs; (3) provision of training manual, handouts, and case studies; (4) highlighting pilot grants to apply newfound knowledge; and (5) facilitating networking opportunities to promote learning networks in which trainees can support each other.

The web-based health portal (Figure 3) was meant to serve as a *one-stop dissemination marketplace*, which would contain

localized health resources to aid health program planners in CBOs to systematically implement evidence-based interventions in their communities. Participants in our project were staff of CBOs in the 3 cities if they met the following criteria [39]: (1) aged ≥18 years, (2) working for CBOs (ie, nonprofit or public service sector organization) in one of the three cities (Boston, Lawrence, and Worcester), and (3) actively involved in developing health promotion programs and efforts in their CBOs.

Figure 3. Screenshot of the web health portal of PLANET MassCONECT.

Project 3: SNAP

Project SNAP was a smartphone-based study that was a collaborative effort between the Dana–Farber Cancer Institute, the University of Saskatchewan, and Baylor College of Medicine and was funded by the Truth Initiative. The objective of the study was to leverage smartphone capabilities in collecting data both passively (eg, collecting geolocation, web browsing, and Bluetooth data) and actively (eg, prompting respondents to answer surveys) to measure exposure to tobacco messages in a real-world setting among participants from underserved groups. Our collaborators from the University of Saskatchewan developed a smartphone app called *Ethica*, which captured complex, rich message exposure data that could use multiple sources to provide information above and beyond what is available in traditional surveys. The app captured data such as (1) responses to ecological momentary assessments in the form of short surveys containing questions on health-information seeking, quit attempts, message exposures, and interpersonal communication regarding tobacco products; (2) photos of pro- or antitobacco messages in the environment; (3) tracking of their web-browsing behaviors; and (4) geolocations.

These three projects were chosen (out of many others in the laboratory) because they each focused on how different forms of eHealth gadgets could close gaps in health disparities. For instance, C2C and PLANET MassCONECT examined the efficacy of using web-based health portals to facilitate health information seeking, whereas SNAP explored the effectiveness of using smartphone apps to capture in situ data (eg, surveys, geolocations, and web search browsing) to understand how external and information environment influence tobacco consumption. Also, these three projects were selected because of the diversity of the categories of participants. For instance, the web-based health portal for C2C was targeted at individuals from underserved groups, whereas the site for PLANET MassCONECT was designed for CBOs that work with underserved groups to find and extract data to promote evidence-based health promotion. SNAP, unlike C2C and PLANET MassCONECT, used smartphone-based methods to examine environmental exposure to tobacco messages among youths of low SEP.

The Five Principles for eHealth Interventions

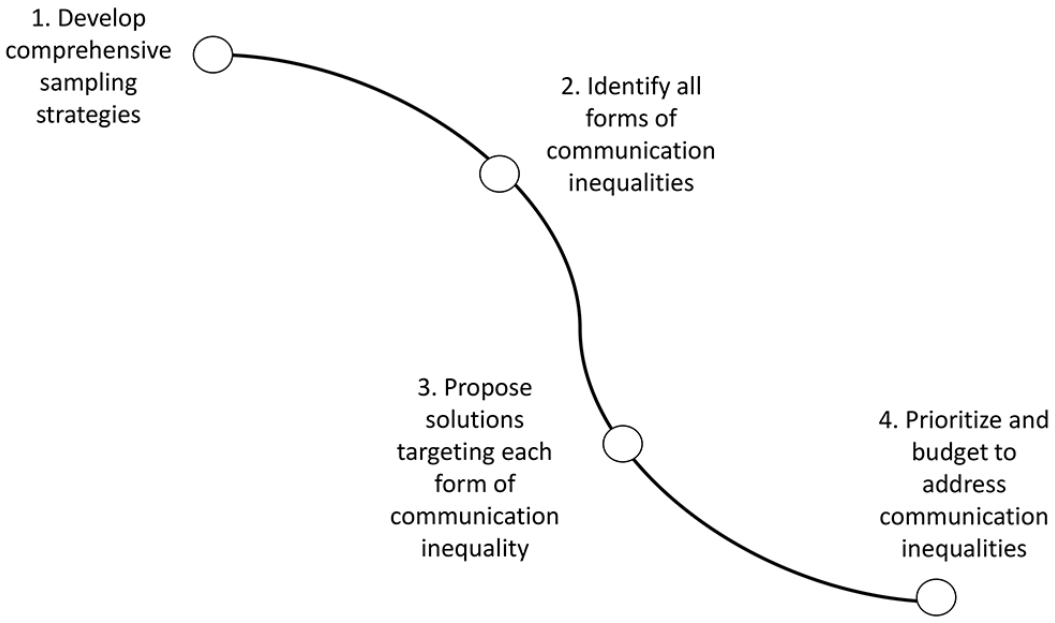
Our experience in working with underserved groups in these digital health projects allows us to draw lessons for effective eHealth intervention research. These lessons were derived primarily from a series of focus groups and usability tests of the eHealth interventions that we conducted for each of the projects, as well as from feedback from community partners. Here, we condensed and extracted five principles: (1) develop a strategic road map to address communication inequalities, (2) engage multiple stakeholders from the beginning for the long haul, (3) design with usability in mind,—enhancing readability and navigability—(4) build privacy safeguards into eHealth interventions and communicate privacy–utility tradeoffs that

come with simplicity, and (5) strive for an optimal balance between open science aspirations and protection of underserved groups. Multimedia Appendix 1 summarizes the objectives of each project, the target groups, and the lessons and principles learned.

Principle 1: Develop a Strategic Road Map to Address Communication Inequalities

First, one of the most important lessons is that it is crucial for research teams to intentionally put in place a strategic road map detailing specific approaches to reduce the burden posed by different forms of *communication inequalities* even before designing eHealth interventions. The presence of communication inequalities poses significant barriers to the underserved population regarding taking advantage of health benefits from using eHealth interventions [40]. Decades of research have documented that even when presented with the same technologies or health information, the ones that benefited the most were people from higher SEPs, whereas people from lower socioeconomic groups were consistently left behind. It goes to show that no matter how sophisticated or cutting edge the technologies are, eHealth interventions are not magic bullets. Instead, introducing eHealth interventions without understanding the broader context of underserved groups will exacerbate the inequalities and preclude people from lower socioeconomic groups from taking advantage of health technologies. As such, without paying attention to what these communication inequalities are, implementing eHealth interventions may inadvertently widen gaps in health inequalities. Although the idea of a strategic roadmap in itself is not novel, there are very few that have explicitly championed the need for a roadmap when working with the underserved [7]. Therefore, what should go into the strategic roadmap? We highlight four key components (Figure 4) to (1) develop comprehensive sampling strategies; (2) identify all forms of communication inequalities, including contextual conditions, and understand the digital media landscape of underserved populations; (3) propose solutions targeting each form of communication inequality; and (4) prioritize and budget for alleviation measures.

Figure 4. Strategic road map for addressing communication inequalities.



First, researchers have to be cognizant of the fact that underserved groups may be either reluctant to participate or face barriers to participation in research studies. For instance, those from racial minorities may be wary of participation in experiments, given the poor experience and exploitation in the past, such as in the infamous Tuskegee experiment in which researchers withheld penicillin for the treatment of their participants' syphilis, although the treatment was widely available [41]. Alternatively, day-to-day experiences with discrimination and poor treatment from those in power may color their experience regarding research. Several reasons have been given for this, such as involved participation risks (eg, participants may not want to identify with researchers' labeling of fear of harm) and opportunity costs (eg, participating in research might mean less time for work or attending doctors' appointment) [42]. To address these concerns, researchers should schedule in-person face-to-face contact with community leaders and organizations, as well as recruitment presentations and meetings in the community where they could present the objectives of the study, articulate how participation in research would benefit the community directly, and address queries and for the community to know the researchers in person.

Second, the road map should identify different forms of communication inequalities. Perhaps the most visible form of communication inequality would be the *lack of access* to these technologies, be it desktops, laptops, smartphones, or wearables

[43], in addition to other factors, such as literacy level and the degree of confidence and efficacy in navigating digital media. Requiring participants from underserved groups to take a portion of their paycheck to purchase communication gadgets is nearly unthinkable, considering their financial constraints. Although this is intuitive for the most part and for many researchers working with urban or rural people from lower socioeconomic groups, what is less apparent is factoring in the hidden costs that underserved populations need to pay to maintain the connection and use of the device. This is known as *connection maintenance cost*, or recurring expenditure, which could refer to the additional financial, time, and energy resources that people from lower socioeconomic groups need to pay to use eHealth interventions [12,17]. For instance, people from underserved groups are more likely to face difficulties in dealing with technical difficulties and experience intermittent internet connection [44].

After identifying these factors, researchers need to propose tangible solutions to address each of these communication inequalities and reflect their commitment through their research budgets. For C2C and SNAP, we provided not only the actual devices required for interventions (eg, computers, web portals, and smartphone apps) but also paid for broadband internet access for participants to ensure that they do not face intermittent connectivity problems, which would dilute the effectiveness of the interventions. We saw some degree of success, especially for our C2C project, where we found that when provided with internet access, participants from underserved groups increased the use of the internet for capital-enhancing purposes, such as seeking health information, contrary to critics who assumed that underserved groups would only use the internet for entertainment purposes [45].

In addition, beyond identifying *deficits* in access to communication technologies, research teams and researchers could creatively design eHealth interventions by leveraging the communication gadgets that the underserved groups are already using instead of introducing or imposing newer forms of technologies that might be met with skepticism. For instance, instead of asking participants to purchase wearables for health tracking, which would be out of their budget, research teams could design health apps that achieve the same functions but at little to no cost to participants. For this reason, for our SNAP project, the Ethica app was repurposed to allow our participants to engage in participatory research on tracking exposure to tobacco messages in their external and web-based environment at no cost.

Principle 2: Engage Multiple Stakeholders From the Beginning for the Long Haul

In the pursuit of developing technologies for underserved populations, the temptation is to develop a technological *messiah complex*, believing that the success and failure of eHealth interventions in improving health rest solely on researchers or the technology. Our experience in working with underserved populations through the years has demonstrated that even well-resourced institutions can be rendered ineffective if they adopt a lone wolf approach in the implementation of eHealth interventions among people from lower socioeconomic groups. Even large anchor institutions such as hospitals need to partner with both public and private organizations if they want to make a sustainable change in the community in which they are located [46], as they may not have a comprehensive understanding of the challenges faced by people in their neighborhood. As such, we advocate researchers to consider a participatory approach and involve and engage different stakeholders from the get-go when designing technology.

We found that engaging multiple stakeholders from the start is nonnegotiable in the development and implementation of eHealth interventions. Although researchers may possess the technological expertise (eg, developing web portals and programming apps for mobile phone solutions) and rigorous training in social-scientific theories, they do not have expertise on the contextual conditions in which the interventions are implemented. Such expertise lies with the users and stakeholders in the community. Depending on the context, there may be a lack of trust in the researchers (ie, underserved populations may have fears about how their data may be misused), uncertainty in how the results would be used (ie, profiling of racial minorities), or a lack of awareness and education regarding the benefits of incorporating the interventions [47]. Therefore, by identifying key stakeholders and forging long-term synergistic collaborations, researchers could benefit from stakeholders' connections within underserved groups and provide scientific expertise that would be relevant in equipping and empowering stakeholders.

So, who are the key stakeholders, and what would be the best way to engage them? There are many types of organizations and individuals that can directly or indirectly influence the success or failure of eHealth interventions [20,48]. Here, we highlight three distinct groups: (1) CBOs, (2) individuals from underserved groups, and (3) their social networks. CBOs play an important public health function within underserved groups. They are well-entrenched within the communities they work with and thus would have greater credibility [49,50], a better grasp on the daily health and financial challenges faced by individuals in the communities, and an intuitive sense of what type of interventions would work. For individuals, it is critical that researchers start to think of the target audience of our eHealth interventions not as passive receivers of information and help but as active members of the communities they are in. Beyond reaching a specific individual, it is critical to engage people as a group, as research has shown that social capital and interpersonal communication can facilitate information sharing and trust and leverage collective actions to foster a sense of community when using eHealth interventions [27,29].

For PLANET MassCONECT, our strategy was to add value to CBOs by developing a web-based health portal that could function as a one-stop-shop by pulling public health data that would be relevant for their day-to-day work. However, we wanted to do so in a way that would organically foster a culture of learning that would be sustainable and continue even after our prescribed intervention period. To do so, we intentionally structured the web-based portal training to facilitate the fostering of social networks and strengthening of ties and relationships among staff members from different CBOs. Apart from the staff

members receiving the most up-to-date knowledge and training in the use of web-based portals to enhance their work, the most valuable work was the creation of a network of dissemination specialists [39] where they could lean on each other for resources when the project ended.

For project SNAP, we attempted to leverage the ubiquity of mobile phones by promoting *citizen science* among our participants. Citizen science is a participatory approach where people are actively engaged in data collection and knowledge translation [51], as opposed to *subjects* where their role would be more passive. To understand the degree of disproportionate tobacco advertising targeted at underserved groups in Boston and Houston, we empowered our participants to take on the role of a collaborator in data collection, where they would snap photos of tobacco messages they encountered in their built environment for a period of 8 weeks. We also monitored their social interaction and how often they smoked together and engaged in conversations regarding smoking and quit attempts.

Although we advocate long-term collaborations with multiple stakeholders, we are mindful that the downsides of collaborations are that the formation of new relationships is resource intensive, and some stakeholders may ultimately not fulfill their obligations. To circumvent these problems, existing implementation science and public health research have advocated having the following three components to ensure that community engagement projects are sustainable over the long run: (1) institutionalized participation, (2) investment in communities, and (3) knowledge production and transfer.

In the context of eHealth intervention projects, although developing and managing relationships with stakeholders could be challenging, a structured and systematic approach to institutionalizing the participation of different stakeholders through formal agreements and engagement plans could provide a scaffold to support such collaborations. Moving beyond formal agreements, investing in communities that health organizations serve could foster greater trust among different stakeholders. These could be achieved by building capacity among different stakeholders, raising digital skills, creating collaborative networks and sharing resources, and empowering individuals. An example of this could be large anchor organizations (eg, large hospitals in communities) taking the lead and creating communities of practice, where other health organizations can come together periodically (eg, annually) to share best practices of using eHealth interventions to address health disparities and their challenges [17]. Finally, the knowledge gained from eHealth intervention projects should be translated into actual practice, and that is where smaller organizations with experience working with underserved groups could provide contextual insights on what works and under what circumstances.

Principle 3: Design With Usability in Mind: Enhancing Readability and Navigability

One of the most difficult challenges in undertaking eHealth intervention design for underserved groups is that besides tackling issues pertaining to technology, design, or even costs, public health organizations, researchers, and technology developers need to understand how usable technologies are to underserved populations in their day-to-day lives. One of the pitfalls for public health researchers involved in building and leveraging communication technologies is having the implicit belief that having technology itself accompanied by the big data it collects as well as sophisticated artificial intelligence algorithms to analyze it would be the panacea in solving health problems brought about by disparities [52]. However, our experiences informed us that at the end of the day, usability is a major factor in technology adoption [53].

When we conducted usability testing for our C2C web portal, we found that one aspect of usability that was often overlooked was the *readability* of the information presented. We found that although individuals from underserved groups expressed interest in using health web portals, one of the main hindrances was site *readability*—whether the content is easily digestible by the intended participants. The participants had very low literacy skills and did not understand some of the languages presented on various pages of the site, although it was perceived as *simple* for the researchers. It became apparent that for the web portal intervention to work, the language would need to be simplified to a very basic level, and terms should reflect literal meanings (eg, no difficult words or metaphors). In addition, even when we provided links to external resources, these websites should be curated for the suitability of low literacy audiences. The focus on content and designing the portal, with readability as a priority, was a crucial factor determining its success through three pathways. First, it has been well established by research that when individuals pay attention to media content [30,54], they are more likely to gain knowledge. Second, research has shown that when individuals pay attention to content delivered via the internet (ie, web portals), they are more likely to engage with others in interpersonal discussion [30], eventually leading to the acquisition of health knowledge, as individuals within social networks help make sense and connect the dots. Third, websites may motivate elaborative processing, the act of connecting new pieces of information with one's knowledge, associated with knowledge gains [31].

A second aspect of usability that researchers should pay attention to is the *navigability* aspect—the ease of traversing one page to another to look for information or activate certain functions. In our C2C web portal (Figure 2), we deliberately designed the main navigation panel to be situated on the left-hand side of the webpage without any top panels that one might see in other similar sites. In the original design, we had a top panel with links to our health content for participants and a left panel with links to external sites, similar to existing health portals. However, the participants were confused as to why there were 2 panels and often relied on the left panel bar, which only brought them to external sites, and they missed out on accessing the actual health content we had prepared for them, which was accessible only from the top panel. To achieve a parsimonious look and simplify use for our participants, we consolidated the navigation panel to the left side of the portal.

For the SNAP project, we found that it is critical to communicate successful navigation with clear visual cues when designing mobile apps for health, although it may be simple. One of the key features of the app was enabling participants to take photos

of anti- or pro-tobacco messages they came across in their daily lives (eg, billboards, advertisements, and shops), which allowed us to map out hotspots of tobacco messaging [31]. In the earlier version of the app, participants needed to manually return to the home page of the app after submitting the photo. However, as they were not brought back to the main page directly, they were confused about whether their photo submissions were successful. Another key feature of our app was a snooze function that enabled participants to pause passive data collection by the app should they want privacy at any point in time. However, it was not clear to the participants if they were successful in pausing the data collection after hitting the snooze function, and some went to their phone settings to manually disable the geolocation tracking by the app. After the feedback, we designed a pop-out SMS text message that clearly highlighted that data collection was paused after they clicked the snooze function.

On the basis of our experiences, we list the following concrete steps that researchers could consider when implementing eHealth interventions in low-resource settings to enhance readability and navigability. First, we strongly recommend that usability testing sessions be designed to explicitly obtain data on how easy it is for participants from underserved groups to comprehend the health content administered through digital gadgets, as well as asking them to provide comments and feedback on how easy (or difficult) it is to navigate and use it. Second, in addition to structuring the usability sessions in a traditionally *passive* manner where participants simply perform the tasks required, researchers could adopt a *think-aloud* protocol of interviewing [55]. The think-aloud interview requires participants to engage in certain tasks: it could be reading content or navigating features of certain health technology and simultaneously expressing their thoughts on whether they found the tasks easy or difficult, which would be recorded [56]. Then, the data could be transcribed and analyzed quantitatively for all participants. In this way, gaps in readability and navigability could be systematically captured and analyzed, providing researchers with empirical insights into areas where significant changes should be made.

Principle 4: Build Privacy Safeguards Into eHealth Interventions and Communicate Privacy-Utility Tradeoffs in Simplicity

With the increasing amount of data collected through eHealth interventions and the public scandals around privacy intrusion and data breaches, it is critical that researchers make privacy and ethics a conversation to have with participants right from the get-go. As the smartphone app used in our SNAP project aimed to collect high-density spatiotemporal data from participants ranging from their web searches to geolocations and the social networks they came into contact with, which are highly personal, we aimed to communicate privacy protection measures to our participants even at the app design stage.

There were several levels of privacy safeguards that were put into place. On a macro level, the app was compliant with the General Data Protection Regulation requirements [52,57], where participants had the right to access or delete their data. In addition, the app was programmed to collect data passively as participants went about their daily lives with a snooze function that they could turn on if they wanted to stop data collection temporarily for any reason.

Although these safeguards were in place, we found that the biggest gap was in the area of communication of privacy measures. During our usability testing, many participants did not understand the function of the snooze button and why it was necessary; many, in fact, suggested removing the snooze function altogether. This highlighted a difference in how participants and researchers approached issues of privacy and indicated to us that some participants might not fully grasp concepts of data privacy, protection, and rights and ownership that we as eHealth interventions researchers are so used to. As such, we recommend that researchers intentionally involve participants from underserved groups in designing privacy features through either focus group or usability testing sessions. This would allow researchers to hear the participants' concerns. In other words, researchers should dedicate a portion of the eHealth intervention training to explain clearly what privacy is, why it matters to the participants, what data are collected, what are the rights over their data, and how they could exercise those rights.

Principle 5: Strive for an Optimal Balance Between Open Science Aspirations and Protection of Underserved Groups

Finally, in the field of eHealth interventions, it is inevitable for researchers to engage in the conversation between the need to strive for *open science* and protect the voluminous amount of data obtained from communication technologies used in the interventions. The push for researchers to adopt the open science movement is a fairly recent phenomenon, driven in part by the rise in big data and artificial intelligence research [58]. Researchers who subscribe to the open science movement aim to foster collaborative networks across research institutions and countries by promoting a culture of *openness* in sharing data and algorithms alongside publications. There are many advantages to the open science paradigm, which address the replicability crisis in social science by enabling other researchers to reproduce similar results with the same data and code, prevent questionable research practices such as p-hacking, and promote best practices of code writing and analysis where others could see and provide constructive criticisms on best practices for data preparation as well as code writing [59].

For many researchers who work in the intersection between eHealth interventions and health promotion among underserved populations, such as ours, we find ourselves in an ideological conundrum. As social scientists, we aspire toward the rigor and collaborative culture of the open science paradigm; on the other hand, we are bound by our commitment to protectively guarding the data collected from underserved groups, knowing that the data could further penalize them if they fall into the wrong hands. For instance, we could track the websites that our C2C participants visited, although they may not be directly related to the project. Through our SNAP project, we collected data at

a granular level where we could know and map individuals' web-browsing patterns, know where they travel to (down to a specific latitude and longitude), and gain insights into the frequency of social contact with their friends captured by Bluetooth data, which provided information when our participants were physically co-located in a given environment. If malicious actors or organizations gain access to these data, they could capitalize on the insights drawn from the data at the expense of people from lower socioeconomic groups.

Although we as a community are still learning to balance this tension, there are several practical steps that one could take to achieve a reasonable balance between embracing open science and data protection. The most basic step researchers could take is to always aggregate data so that sensitive details of individuals would retain a higher level of anonymity. For our C2C intervention, although we could potentially track the web-browsing behaviors of underserved populations, we chose to collect data at the household (ie, aggregate) level, knowing that this tradeoff would provide an additional layer of protection for our participants. As for geolocations captured through smartphones in our SNAP study, although we could analyze the data at the most granular level (eg, the exact locations they have visited down to the seconds), we collapsed the geolocations into 10-minute bins for analyses, making it potentially more difficult to trace the geolocation patterns to specific individuals. Analyzing data at the aggregated level gave us the confidence to present sensitive information (eg, hotspot maps) without compromising the privacy of our participants.

A second practical way to balance the tension between open science and participants' privacy is to host the data and analyze them through a secure remote data storage and analysis environment as much as possible, as opposed to storing and analyzing data in researchers' desktops in the office or personal laptops. This is part of good data hygiene; after all, the security of personal desktops and laptops has a higher likelihood of being compromised compared with a secure cluster computing environment. For instance, we relied on the remote computing environment provided by the Institute of Quantitative Social Sciences at Harvard University for data management and analysis of part of our SNAP data. Apart from the universities' cluster computing environment, researchers may seek out trustworthy companies involved in cloud computing to curate their data management and analysis pipeline.

Third, researchers could promote or engage in *selective openness* depending on the nature of their eHealth interventions. For our capacity-building work with CBOs in Planet MassCONECT, the sharing of data and best practices were done in a safe zone, where the audiences are staff members from different CBOs involved in similar work with underserved groups. Thus, openness is achieved in an environment of trust and familiarity, where information sharing on best practices would benefit all CBOs. At the same time, by only sharing data and information with those invested in the communities, it prevents external organizations or individuals from accessing sensitive data.

Limitations

We are mindful that while we aim to provide a synthesis of the principles over the decade, there are several limitations. First, we recognize that every research team and organization working with underserved groups have unique challenges and that some of our recommendations may not be directly applicable. As such, it is important for research teams to be keenly aware of their own unique contexts and only apply those strategies and principles that are most relevant. Second, as most of our work was conducted in the United States, there may be cultural barriers specific to different countries. For instance, cultural worldviews such as fatalism may further impede eHealth intervention adoption because of the inherent belief that there was nothing that patients could do to avoid contracting certain diseases [29,60].

How the Proposed Principles Complement Existing Knowledge in eHealth Interventions

Despite the limitations, there are several ways in which the proposed principles will address gaps in knowledge within existing eHealth intervention research. Currently, research on eHealth interventions in the context of health disparities typically focuses on improving usability design [61], identifying factors to improve adoption [3], and how digital health technologies may amplify inequality [6]. We build upon the current knowledge of eHealth interventions in a few ways. First, we argue that the success of eHealth interventions goes beyond usability and interface issues, and researchers and technology developers need to pay attention to the broader context that would influence the eventual adoption and use of digital health technologies. Second, we strongly advocate that eHealth interventions will need to be embraced and supported by multiple stakeholders over the long run to have a tangible impact on the health of underserved groups. This would involve building networks and technological capacity among under-resourced CBOs so that they could effectively leverage eHealth interventions in their daily work. Finally, we highlight the need for the research community to be transparent and address privacy concern issues collectively with participants from the get-go to balance the benefits of open science and protect the data of communities that embrace the use of these technologies. Future research could extend our research by studying the broader contextual factors that facilitate or impede the use of eHealth interventions by health researchers using a case study approach [62]. This would enable researchers to gain a holistic understanding of the specific opportunities and constraints faced by organizations in eHealth intervention research. In addition, future scholars could conduct longitudinal field experiments to ascertain the effectiveness of eHealth interventions in addressing health disparities over the long run, as well as to systematically examine how openness and transparency with privacy issues may improve trust and adoption of eHealth interventions among underserved groups.

Conclusions

There has never been a more exciting time to be involved in the development and implementation of eHealth interventions among underserved groups. With advancements in big data platforms and artificial intelligence, there are multiple opportunities to leverage technologies and data to improve health for the underserved [17]. However, the aspects of technology design and data analysis are small puzzles in the scheme of the big picture. The implementation of successful eHealth interventions ultimately rests on how well researchers understand the barriers to and facilitators of technology acceptance. Beyond that, researchers need to understand the digital media landscape of underserved populations, navigate and grasp sociocultural norms, as well as bring organizations, communities, and individuals on the same page to address health disparities through technology.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Summary of eHealth intervention projects in the Viswanath laboratory.
[DOCX File , 18 KB - jmir_v24i1e25419_app1.docx]

References

1. Reen GK, Muirhead L, Langdon DW. Usability of health information websites designed for adolescents: systematic review, neurodevelopmental model, and design brief. J Med Internet Res 2019 Apr 23;21(4):e11584 [FREE Full text] [doi: 10.2196/11584] [Medline: 31012856]
2. Tanaka R, Banerjee A, Surikova J, Tracey J, Payne A, Ross H, et al. A moderated e-forum for adults with cardiovascular disease: usability study. JMIR Hum Factors 2018 May 18;5(2):e20 [FREE Full text] [doi: 10.2196/humanfactors.8820] [Medline: 29776901]
3. Grossman L, Creber RM, Benda NC, Wright D, Vawdrey DK, Ancker JS. Interventions to increase patient portal use in vulnerable populations: a systematic review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):855-870 [FREE Full text] [doi: 10.1093/jamia/ocz023] [Medline: 30958532]
4. Simons D, De Bourdeaudhuij I, Clarys P, De Cocker K, Vandelanotte C, Deforche B. A smartphone app to promote an active lifestyle in lower-educated working young adults: development, usability, acceptability, and feasibility study. JMIR Mhealth Uhealth 2018 Feb 20;6(2):e44 [FREE Full text] [doi: 10.2196/mhealth.8287] [Medline: 29463491]
5. Torous J, Rodriguez J, Powell A. The new digital divide for digital biomarkers. Digit Biomark 2017 Sep;1(1):87-91 [FREE Full text] [doi: 10.1159/000477382] [Medline: 29104959]
6. Veinot T, Mitchell H, Ancker JS. Good intentions are not enough: how informatics interventions can worsen inequality. J Am Med Inform Assoc 2018 Aug 01;25(8):1080-1088 [FREE Full text] [doi: 10.1093/jamia/ocy052] [Medline: 29788380]
7. Viswanath K. Public communications and its role in reducing and eliminating health disparities. In: Thomson GE, Mitchell F, Williams MB, editors. Examining the Health Disparities Research Plan of the National Institutes of Health: Unfinished Business?. Washington, DC: The National Academies Press; 2006:215-253.
8. Weiss D, Rydland HT, Øversveen E, Jensen MR, Solhaug S, Krokstad S. Innovative technologies and social inequalities in health: a scoping review of the literature. PLoS One 2018 Apr 3;13(4):e0195447 [FREE Full text] [doi: 10.1371/journal.pone.0195447] [Medline: 29614114]
9. Peacock S, Reddy A, Leveille SG, Walker J, Payne TH, Oster NV, et al. Patient portals and personal health information online: perception, access, and use by US adults. J Am Med Inform Assoc 2017 Apr 01;24(e1):173-177. [doi: 10.1093/jamia/ocw095] [Medline: 27413120]
10. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018 May 8;1(1):18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
11. Zhang X, Pérez-Stable EJ, Bourne PE, Peprah E, Duru OK, Breen N, et al. Big data science: opportunities and challenges to address minority health and health disparities in the 21st century. Ethn Dis 2017 Apr 20;27(2):95-106 [FREE Full text] [doi: 10.18865/ed.27.2.95] [Medline: 28439179]
12. Viswanath K, Nagler RH, Bigman-Galimore CA, McCauley MP, Jung M, Ramanadhan S. The communications revolution and health inequalities in the 21st century: implications for cancer control. Cancer Epidemiol Biomarkers Prev 2012 Oct 07;21(10):1701-1708. [doi: 10.1158/1055-9965.epi-12-0852]
13. Antonio M, Petrovskaya O, Lau F. Is research on patient portals attuned to health equity? A scoping review. J Am Med Inform Assoc 2019 Aug 01;26(8-9):871-883 [FREE Full text] [doi: 10.1093/jamia/ocz054] [Medline: 31066893]
14. Sim D, Lim K. Coronavirus: why aren't Singapore residents using the TraceTogether contact-tracing app? South China Morning Post. 2020. URL: https://www.scmp.com/https://www.scmp.com/week-asia/people/article/3084903/coronavirus-why-arent-singapore-residents-using-tracetogether [accessed 2020-05-18]

15. Schaefer GO, Ballantyne A. Downloading COVID-19 contact tracing apps is a moral obligation. BMJ Blogs. 2020. URL: https://blogs.bmj.com/medical-ethics/2020/05/04/downloading-covid-19-contact-tracing-apps-is-a-moral-obligation/ [accessed 2021-11-16]
16. Viswanath K, Lee EW, Pinnamaneni R. We need the lens of equity in COVID-19 communication. Health Commun 2020 Dec 26;35(14):1743-1746. [doi: 10.1080/10410236.2020.1837445] [Medline: 33106029]
17. Lee EW, Viswanath K. Big data in context: addressing the twin perils of data absenteeism and chauvinism in the context of health disparities research. J Med Internet Res 2020 Jan 07;22(1):e16377 [FREE Full text] [doi: 10.2196/16377] [Medline: 31909724]
18. Kontos EZ, Emmons KM, Puleo E, Viswanath K. Communication inequalities and public health implications of adult social networking site use in the United States. J Health Commun 2010 Dec;15 Suppl 3:216-235 [FREE Full text] [doi: 10.1080/10810730.2010.522689] [Medline: 21154095]
19. McLeroy KR, Bibeau D, Steckler A, Glanz K. An ecological perspective on health promotion programs. Health Educ Q 1988 Sep 04;15(4):351-377. [doi: 10.1177/109019818801500401] [Medline: 3068205]
20. Moran MB, Frank LB, Zhao N, Gonzalez C, Thainiyom P, Murphy ST, et al. An argument for ecological research and intervention in health communication. J Health Commun 2016 Jan;21(2):135-138 [FREE Full text] [doi: 10.1080/10810730.2015.1128021] [Medline: 26788682]
21. Viswanath K, McCloud RF, Bekalu MA. Communication, health and equity: structural influences. In: Thompson TL, Harrington NG, editors. The Routledge Handbook of Health Communication (3rd edition). New York: Routedge; 2021.
22. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Beyond access: barriers to internet health information seeking among the urban poor. J Am Med Inform Assoc 2016 Nov;23(6):1053-1059 [FREE Full text] [doi: 10.1093/jamia/ocv204] [Medline: 27206459]
23. Viswanath K, Kreuter MW. Health disparities, communication inequalities, and eHealth. Am J Prev Med 2007 May;32(5 Suppl):131-133 [FREE Full text] [doi: 10.1016/j.amepre.2007.02.012] [Medline: 17466818]
24. Gonzales AL. Health benefits and barriers to cell phone use in low-income urban U.S. neighborhoods: indications of technology maintenance. Mobile Media Commun 2014 Aug 19;2(3):233-248. [doi: 10.1177/2050157914530297]
25. Smith SG, O'Conor R, Aitken W, Curtis LM, Wolf MS, Goel MS. Disparities in registration and use of an online patient portal among older adults: findings from the LitCog cohort. J Am Med Inform Assoc 2015 Jul;22(4):888-895 [FREE Full text] [doi: 10.1093/jamia/ocv025] [Medline: 25914099]
26. Sarkar U, Karter AJ, Liu JY, Adler NE, Nguyen R, López A, et al. Social disparities in internet patient portal use in diabetes: evidence that the digital divide extends beyond access. J Am Med Inform Assoc 2011 May 1;18(3):318-321 [FREE Full text] [doi: 10.1136/jamia.2010.006015] [Medline: 21262921]
27. Ackerson LK, Viswanath K. The social context of interpersonal communication and health. J Health Commun 2009 May 20;14 Suppl 1(sup1):5-17. [doi: 10.1080/10810730902806836] [Medline: 19449264]
28. de Souza R. Creating "communicative spaces": a case of NGO community organizing for HIV/AIDS prevention. Health Commun 2009 Dec 10;24(8):692-702. [doi: 10.1080/10410230903264006] [Medline: 20183378]
29. Lee EW, Ho SS, Chow JK, Wu YY, Yang Z. Communication and knowledge as motivators: understanding Singaporean women's perceived risks of breast cancer and intentions to engage in preventive measures. J Risk Res 2013 Aug;16(7):879-902. [doi: 10.1080/13669877.2012.761264]
30. Lee EW, Shin M, Kawaja A, Ho SS. The augmented cognitive mediation model: examining antecedents of factual and structural breast cancer knowledge among Singaporean women. J Health Commun 2016 Dec;21(5):583-592. [doi: 10.1080/10810730.2015.1114053] [Medline: 27128006]
31. Lee EW, Ho SS. Staying abreast of breast cancer: examining how communication and motivation relate to Singaporean women's breast cancer knowledge. Asian J Commun 2015 Feb 19;25(4):422-442. [doi: 10.1080/01292986.2014.976580]
32. Ramanadhan S, "Vish" Viswanath K. Engaging communities to improve health: models, evidence, and the participatory knowledge translation (PaKT) framework. In: Principles and Concepts of Behavioral Medicine. New York: Springer; 2018:679-711.
33. McCloud RF, Bekalu MA, Maddox N, Minsky SJ, Viswanath K. Leveraging breadth and depth: strategies to characterize population diversity to address cancer disparities in the DF/HCC catchment area. Cancer Epidemiol Biomarkers Prev 2018 Oct 19;28(3):435-441. [doi: 10.1158/1055-9965.epi-18-0799]
34. Viswanath K, McCloud R, Minsky S, Puleo E, Kontos E, Bigman-Galimore C, et al. Internet use, browsing, and the urban poor: implications for cancer control. J Natl Cancer Inst Monogr 2013 Dec;2013(47):199-205 [FREE Full text] [doi: 10.1093/jncimonographs/lgt029] [Medline: 24395992]
35. Piwek L, Ellis DA, Andrews S, Joinson A. The rise of consumer health wearables: promises and barriers. PLoS Med 2016 Feb;13(2):e1001953 [FREE Full text] [doi: 10.1371/journal.pmed.1001953] [Medline: 26836780]
36. Bekalu MA, Minsky S, Viswanath K. Beliefs about smoking-related lung cancer risk among low socioeconomic individuals: the role of smoking experience and interpersonal communication. Glob Health Promot 2019 Sep 07;26(3):88-93. [doi: 10.1177/1757975917732758] [Medline: 29110569]

37. Nagler RH, Ramanadhan S, Minsky S, Viswanath K. Recruitment and retention for community-based eHealth interventions with populations of low socioeconomic position: strategies and challenges. J Commun 2013 Feb 01;63(1):201-220 [FREE Full text] [doi: 10.1111/jcom.12008] [Medline: 23439871]
38. Ramanadhan S, Mendez SR, Rao M, Viswanath K. Social media use by community-based organizations conducting health promotion: a content analysis. BMC Public Health 2013 Dec 05;13:1129 [FREE Full text] [doi: 10.1186/1471-2458-13-1129] [Medline: 24313999]
39. Ramanadhan S, Minsky S, Martinez-Dominguez V, Viswanath K. Building practitioner networks to support dissemination and implementation of evidence-based programs in community settings. Transl Behav Med 2017 Sep 5;7(3):532-541 [FREE Full text] [doi: 10.1007/s13142-017-0488-y] [Medline: 28382582]
40. Viswanath K, Ramanadhan S, Kontos EZ. Mass media. In: Galea S, editor. Macrosocial Determinants of Population Health. New York: Springer; 2007:275-294.
41. U.S. public health service syphilis study at Tuskegee: the Tuskegee timeline. Centers for Disease Control and Prevention. 2020. URL: https://www.cdc.gov/tuskegee/timeline.htm [accessed 2021-11-16]
42. Ellard-Gray A, Jeffrey NK, Choubak M, Crann SE. Finding the hidden participant. Int J Qual Methods 2015 Dec 17;14(5):160940691562142. [doi: 10.1177/1609406915621420]
43. Granja C, Janssen W, Johansen MA. Factors determining the success and failure of eHealth interventions: systematic review of the literature. J Med Internet Res 2018 May 01;20(5):e10235 [FREE Full text] [doi: 10.2196/10235] [Medline: 29716883]
44. Gonzales AL, Ems L, Suri VR. Cell phone disconnection disrupts access to healthcare and health resources: a technology maintenance perspective. New Media Society 2016 Jul 10;18(8):1422-1438. [doi: 10.1177/1461444814558670]
45. McCloud RF, Okechukwu CA, Sorensen G, Viswanath K. Entertainment or health? Exploring the internet usage patterns of the urban poor: a secondary analysis of a randomized controlled trial. J Med Internet Res 2016 Mar 03;18(3):e46 [FREE Full text] [doi: 10.2196/jmir.4375] [Medline: 26940637]
46. Vize R. Hospitals as anchor institutions: how the NHS can act beyond healthcare to support communities. Br Med J 2018 May 14:k2101. [doi: 10.1136/bmj.k2101]
47. Bonevski B, Randell M, Paul C, Chapman K, Twyman L, Bryant J, et al. Reaching the hard-to-reach: a systematic review of strategies for improving health and medical research with socially disadvantaged groups. BMC Med Res Methodol 2014 Mar 25;14(1):42 [FREE Full text] [doi: 10.1186/1471-2288-14-42] [Medline: 24669751]
48. Lee EW, Zheng H, Aung HH, Seidmann V, Li C, Aroor MR, et al. Examining organizational, cultural, and individual-level factors related to workplace safety and health: a systematic review and metric analysis. Health Commun 2021 May 09;36(5):529-539. [doi: 10.1080/10410236.2020.1731913] [Medline: 32146838]
49. Ramanadhan S, Crisostomo J, Alexander-Molloy J, Gandelman E, Grullon M, Lora V, PLANET MassCONECT C-PAC, et al. Perceptions of evidence-based programs among community-based organizations tackling health disparities: a qualitative study. Health Educ Res 2012 Aug 19;27(4):717-728 [FREE Full text] [doi: 10.1093/her/cyr088] [Medline: 22012968]
50. Stephens KK, Rimal RN, Flora JA. Expanding the reach of health campaigns: community organizations as meta-channels for the dissemination of health information. J Health Commun 2004 Jan;9 Suppl 1(sup1):97-111. [doi: 10.1080/10810730490271557] [Medline: 14960406]
51. Katapally TR, Bhawra J, Leatherdale ST, Ferguson L, Longo J, Rainham D, et al. The smart study, a mobile health and citizen science methodological platform for active living surveillance, integrated knowledge translation, and policy interventions: longitudinal study. JMIR Public Health Surveill 2018 Mar 27;4(1):e31 [FREE Full text] [doi: 10.2196/publichealth.8953] [Medline: 29588267]
52. Lee EW, Yee AZ. Toward data sense-making in digital health communication research: why theory matters in the age of big data. Front Commun 2020 Feb 27;5:1-10. [doi: 10.3389/fcomm.2020.00011]
53. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
54. Yang X, Chuah AS, Lee EW, Ho SS. Extending the cognitive mediation model: examining factors associated with perceived familiarity and factual knowledge of nanotechnology. Mass Commun Society 2017 Jan 19;20(3):403-426. [doi: 10.1080/15205436.2016.1271436]
55. Lee EW, Bekalu MA, McCloud R, Vallone D, Arya M, Osgood N, et al. The potential of smartphone apps in informing protobacco and antitobacco messaging efforts among underserved communities: longitudinal observational study. J Med Internet Res 2020 Jul 07;22(7):e17451 [FREE Full text] [doi: 10.2196/17451] [Medline: 32673252]
56. Chew S, Lai PS, Ng CJ. Usability and utility of a mobile app to improve medication adherence among ambulatory care patients in Malaysia: qualitative study. JMIR Mhealth Uhealth 2020 Jan 31;8(1):e15146 [FREE Full text] [doi: 10.2196/15146] [Medline: 32003748]
57. Eveland WP, Dunwoody S. Examining information processing on the world wide web using think aloud protocols. Media Psychol 2000 May;2(3):219-244. [doi: 10.1207/s1532785xmep0203_2]
58. Is Ethica GDPR compliant? Ethica. 2020. URL: https://learn.ethicadata.com/faq/is-ethica-gdpr-compliant/ [accessed 2021-11-16]

59. Magalhães J, Hartz Z, Antunes A, Martins MD. An overview of the open science in times of big data and innovation to global health. Int J Innov 2017 Dec 01;5(3):270-288. [doi: 10.5585/iji.v5i3.219]
60. Dienlin T, Johannes N, Bowman ND, Masur PK, Engesser S, Kümpel AS, et al. An agenda for open science in communication. J Commun 2021;71(1):1-26. [doi: 10.1093/joc/jqz052]
61. Leung D, Chow T, Wong E. Cancer-related information seeking and scanning behaviors among older chinese adults: examining the roles of fatalistic beliefs and fear. Geriatrics (Basel) 2017 Dec 03;2(4):38 [FREE Full text] [doi: 10.3390/geriatrics2040038] [Medline: 31011048]
62. Vermeulen J, Neyens J, Spreeuwenberg M, van Rossum E, Sipers W, Habets H, et al. User-centered development and testing of a monitoring system that provides feedback regarding physical functioning to elderly people. Patient Prefer Adher 2013 Aug:843. [doi: 10.2147/ppa.s45897]

Abbreviations

C2C: Click to Connect
CBO: community-based organization
EHR: electronic health record
SEP: socioeconomic position
SNAP: Smartphone App for Public Health

Edited by R Kukafka, G Eysenbach; submitted 10.11.20; peer-reviewed by H Traino, PhD, MPH, B Eapen, I Mircheva, J Lander; comments to author 02.03.21; revised version received 25.06.21; accepted 08.10.21; published 07.01.22.

Please cite as:
Lee EWJ, McCloud RF, Viswanath K
Designing Effective eHealth Interventions for Underserved Groups: Five Lessons From a Decade of eHealth Intervention Design and Deployment
J Med Internet Res 2022;24(1):e25419
URL: https://www.jmir.org/2022/1/e25419
doi:10.2196/25419
PMID:34994700

Viewpoint

The Future of Virtual Care for Older Ethnic Adults Beyond the COVID-19 Pandemic

Quynh Pham[1,2,3*], PhD; Noor El-Dassouki[1*], BSc, MSc; Raima Lohani[1*], BScN, MHI; Aravinth Jebanesan[4], BSc; Karen Young[1,2], BA

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

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

[3]Telfer School of Management, University of Ottawa, Ottawa, ON, Canada

[4]Global Health Office, Faculty of Health Science, McMaster University, Hamilton, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Quynh Pham, PhD
Centre for Global eHealth Innovation
Techna Institute
University Health Network
190 Elizabeth Street
Toronto, ON, M5G 2C4
Canada
Phone: 1 416 340 4800 ext 4765
Email: Q.Pham@uhn.ca

Abstract

The COVID-19 pandemic has fundamentally changed how Canadians access health care. Although it is undeniable that the rapid adoption of virtual care has played a critical role in reducing viral transmission, the gap in equitable access to virtual care remains pervasive for Canada's aging and ethnocultural minority communities. Existing virtual care solutions are designed for the English-speaking, health-literate, and tech-savvy patient population, excluding older ethnic adults who often do not see themselves reflected in these identities. In acknowledging the permanency of virtual care brought on by the pandemic, we have a collective responsibility to co-design new models that serve our older ethnic patients who have been historically marginalized by the status quo. Building on existing foundations of caregiving within ethnocultural minority communities, one viable strategy to realize culturally equitable virtual care may be to engage the highly motivated and skilled family caregivers of older ethnic adults as partners in the technology-mediated management of their chronic disease. The time is now to build a model of shared virtual care that embraces Canada's diverse cultures, while also providing its older ethnic adults with access to health innovations in partnership with equally invested family caregivers who have their health at heart.

(J Med Internet Res 2022;24(1):e29876) doi:10.2196/29876

KEYWORDS

virtual care; digital health; health equity; cultural equity; chronic disease; caregivers; ethnocultural minority; older adults; ethnicity; ethnic patients; technology-mediated care; equity; diversity; family

Virtual Care in a Postpandemic World

The COVID-19 pandemic has fundamentally changed how all Canadians access health care. As initial media coverage showed unsettling scenes of Canada becoming overwhelmed by this new virus, we felt a sense of unease about our own health and the new ways in which we would receive care. With our health care systems stretched and strained to reduce the devastation wreaked by the pandemic, virtual care solutions were rapidly adopted to maintain continuity of care and deliver essential services. Yet in looking deeper at how virtual care became the new standard of care during the pandemic, we know now that this technological transformation did not always benefit those most in need.

Virtual care is defined as "any interaction between patients and/or members of their circle of care, occurring remotely, using any forms of communication or information technologies, with the aim of facilitating or maximizing the quality and effectiveness of patient care" [1]. The first wave of the COVID-19 pandemic forced a sharp decline in in-person

primary care visits in Canada's most populous province of Ontario as phone and video consultations became the de facto means of accessing care [2]. The second wave of the COVID-19 pandemic saw an even greater shift to virtual care delivery across diverse health care settings such as physiotherapy and postoperative care [3]. Although it is undeniable that virtual care has played a critical role in reducing viral transmission, the future of health care cannot simply be a matter of transitioning current models of care to virtual settings. Sustaining virtual care beyond the pandemic requires us to thoughtfully consider the emerging needs of our society. The Canadian population is growing older, and at the same time, more ethnically diverse. When we look at Canada's growing demographic of aging and ethnically diverse adults, the risk of inequity is stark. First, older adults aged 65 years or older comprise 18% of the total population and 9% of the visible minority population [4,5]. Second, visible minority older adults are the fastest growing demographic in Canada [6]. Finally, the impact of chronic disease among older adults places a growing burden on our health care system: over 70% of older adults have been diagnosed with at least one of the 10 most common chronic diseases [7,8]. Without a shared strategy to provide equitable access to health care for this emerging population, virtual care runs the risk of exacerbating existing health care inequities for older ethnic adults living with chronic disease [9,10].

In this viewpoint, we refer to the ethnic groups that make up a minority of the Canadian population according to census data and the cultures which characterize them as "ethnocultural minority" communities [11]. Ethnicity comprises the shared aspects of belonging to a group, such as language, ancestry, nationality, and values [12]. The social practices, behavioral norms, and physical expressions of societies—such as traditions, customs, and ideologies—form their culture [13]. Our use of the term "ethnocultural" captures both the shared group aspects and social practices that define ethnicity and culture respectively. In using this terminology, we aim to highlight the ethnic and cultural aspects of caring as factors that influence how patients are able to access virtual care but are often not considered in the design of virtual care platforms.

Early findings on the spread of virtual care during the pandemic in the United States have revealed that older adults, those with limited English proficiency, and those from ethnocultural minority communities have reduced access to virtual care services [10]. Promising work has been done at the intersection of aging, technology, and diversity to improve equity and access to virtual care; solutions include involving patients in technology co-design, increasing access to language interpreters, developing training materials in collaboration with community-based organizations, equipping community centers with the devices and space to receive virtual care, and offering low-tech care options such as telephone visits to improve access [14,15]. However, existing virtual care solutions are often designed for the English-speaking, health-literate, and tech-savvy patient population, and are informed by a biomedical model of health and illness. Older ethnic adults often do not see themselves reflected in these identities and, as such, continue to be excluded from receiving virtual care. Issuing these communities an ultimatum to successfully learn complex technological and self-management behaviors in order to receive care will likely result in even greater isolation and heighten their existing struggles to safely age in place. This feat can be insurmountable in the best of times and is even more overwhelming during a global pandemic when these communities are already experiencing greater health and economic needs [16-19]. Current virtual care models ultimately fail to acknowledge the reality that the same older ethnic adults who are challenged to adopt virtual care are also more likely to have complex health needs and greater difficulties accessing health care [20]. For a population that already experiences deeply rooted health, social, and economic inequities, the inability to access virtual care nearly guarantees that older ethnic adults will struggle to receive *any* care from increasingly digital-first health systems in a postpandemic world. We believe a new model of virtual care that rejects burdening older ethnic adults with complex requirements in order to access care is desperately needed. One viable strategy to realize culturally equitable virtual care may be to engage the highly motivated and skilled family caregivers of older ethnic adults as partners in the technology-mediated management of their chronic disease.

The Opportunity for Family Caregivers to Become Virtual Care Partners

Virtual care is here to stay. In acknowledging this new reality, we have a rare opportunity to engage older ethnic adults in virtual care by building on the values that are deeply embedded in their collectivist communities [21]. Despite nuances in how caregiving is defined across cultures, there is a common perception that caring for one's family is a natural and expected part of life [22]. Across many Asian cultures, familial kinship and filial piety represent a deep commitment to caring for aging parents [21]. Although familial kinship places greater value on sharing caregiving responsibilities across family members, filial piety focuses specifically on the parent-child relationship and the practice of caring for one's parents [21,23]. Both familial kinship and filial piety emphasize the value of intergenerational relationships and family-centered approaches to caregiving, which see adult children heavily involved in the role of caring for their ailing parents [23]. We can build on the existing foundations of caregiving within ethnocultural minority communities to establish older ethnic adults and their family caregivers as true partners in virtual care.

Acting on the opportunity presented by this model of care requires us to first understand the intrinsic strengths of family caregiving. Family caregivers provide comprehensive and critical assistance to older ethnic adults navigating the health care system [24]. Although this role can be assumed by partners, siblings, grandchildren, extended family, and friends, the vast majority of caregivers in Ontario are adult children [25]. Through a lifetime of shared experiences, adult children bring an invaluable wealth of knowledge about their parents' health, social, economic, and supportive needs. Their linguistic and cultural relation to their parents affords them some jurisdiction to make joint decisions on how to manage health and illness [26]. These values of familial kinship and filial piety, the cultural embeddedness of caregiving, and the ability to inform health

care preferences and practices can empower adult children to not only care for their parents, but also to see the value of participating in chronic disease comanagement using virtual care [21-23,27-31]. Many adult child caregivers are also highly educated and have strong digital and health literacy skills [32-34]. Recent census data examining second-generation Chinese and South Asian Canadians—communities that place high value on filial piety—revealed that these groups are generally highly educated and working in advanced information technology occupations [35]. The adult children of older ethnic patients thus present as an extremely valuable resource with skills that will allow them to be the index user of virtual care technologies in the patient-caregiver partnership, thereby removing the burden on their parents to adopt technology [32].

Leveraging this opportunity to comanage chronic diseases through virtual care also requires us to carefully consider the challenges and implications of family caregiving. Caring for a sick family member is psychologically, physically, socially, and financially demanding; experiences of caregiver burnout are unfortunately common [36-39]. The cost of caring for an older family member and witnessing their loss of independence and functional decline can manifest as negative feelings of loss, hostility, and anxiety among family caregivers and strain relationships [40-42]. Caregivers of ethnocultural minority communities also exhibit higher levels of involvement in care activities compared to other ethnic groups, which can further impede their quality of life [43]. In particular, the cultural norms of many ethnocultural minority communities often place the responsibility of caregiving on wives, daughters, and daughters-in-law [23]. Virtual models of shared care should be designed in a way that reduces this gendered burden and removes the need for women to make personal sacrifices to fulfill their filial responsibilities [44].

Significant privacy implications must be considered when operationalizing a model of shared virtual care. Although there are many benefits to sharing personal health information in this context, patients and their caregivers should be informed of the risks of granting shared access. Having caregivers be privy to previously unknown and oftentimes sensitive information can lead to undesired changes to the parent-child relationship and an unintended loss of patient autonomy [45]. For instance, gaining access to personal information can lead caregivers to dominate care decisions and communication with providers, thereby reducing the patient's participation in, and control of, their own care [46]. Establishing routine communication exercises that revisit roles and responsibilities around shared decision-making and care actions, and creating processes that ensure active patient engagement, can help partners to better navigate their shared roles.

Reimagining Virtual Care to Be Culturally Equitable for Older Ethnic Adults

The future of virtual care for older ethnic adults beyond the COVID-19 pandemic will require a reimagining of the ways in which ethnocultural minority communities access and receive care. Virtual care models that formally engage caregivers have shown promise in improving access to primary and chronic care for older ethnic adults. Latulipe and colleagues [47,48] recently showed that creating caregiver proxy user accounts in patient health portals can allow patients with limited English language proficiency to fully benefit from technologies they would be unable to access otherwise. Piette and colleagues [49-52] have led a decade's worth of formative and definitive trials to study the effectiveness of the CarePartner program and its implementations for heart failure, cancer, diabetes, and mental health. Patients in the CarePartner program complete automated symptom assessments through an interactive voice response (IVR) system and receive tailored health and behavioral advice on how to manage their symptoms. Their caregiver receives a structured report automatically via email with information about the patient's health and what they can do to help. The program has demonstrated significant improvements in symptom management, medication adherence, health service utilization, and emotional coping, with patient-partner dyads reporting improved frequency and quality of communication [51-54]. It is worth noting that the CarePartner Program maintained its effectiveness when deployed in Bolivia to support the dyadic management of diabetes and hypertension [49]. Findings from the randomized trial comparing IVR alone or with automated feedback sent to a caregiver after each IVR assessment showed that caregiver feedback increased engagement with IVR most significantly among patients of Indigenous ethnicity and those with low functional health literacy [55].

Motivated by the successes of dyadic virtual care programs, our team [56] has started to formalize caregiver inclusion in the service design and delivery of our virtual clinics for heart failure and cancer survivorship. Our Medly Clinic [57] supports remote heart failure monitoring [58-61], while our Ned Clinic [62] enables asynchronous prostate cancer survivorship care [63,64]. Both virtual clinics have historically excluded patients who were not comfortable using technology, did not have an email address, and had limited English proficiency; we previously assumed that these factors would challenge patients to use clinic services as intended and derive benefit. The onset of the COVID-19 pandemic created urgency to eliminate barriers to the Medly and Ned clinics as they enabled the provision of services like medication titration and follow-up visits that were paused due to institutional lockdown orders. As a result of increased patient demand for these services, both clinics began enrolling a growing proportion of older ethnic adult patients who were accompanied by their family caregivers and relied on them to adopt the clinic technology. This off-label dyadic approach to onboarding patients who would otherwise not qualify for virtual clinic services brought to light a unique opportunity to formally adapt the clinics for diverse cultural groups. We have since developed Medly Caregiver Live Reports to enable caregivers to view their partner's data generated on the Medly patient app (Figure 1). In parallel to this effort, we drafted new standard operating procedures to formally enroll patient-caregiver dyads into the Ned Clinic and support their shared use of the platform. Our Ned Partners service includes expanded recruitment and onboarding materials, a tips sheet and informal agreement for dyads to align on the challenges of survivorship comanagement, and adapted workflows to accommodate changes in their involvement in the clinic and ownership of care tasks over time (Multimedia Appendix 1).

We believe that Canada's virtual care road map requires systemic reform to improve the health and social outcomes of its ethnocultural minority communities. By 2031, 32% of Canada's population is projected to belong to a visible minority group [65]. As this population ages, they deserve to receive virtual care that they perceive to be acceptable, appropriate, and aligned with their cultural beliefs. We have a collective responsibility to co-design new models of virtual care that reflect these ideals and deliver measurable and sustained impact. Only then can virtual care transition from an exclusive service to an equitable standard that serves those patients who have been historically marginalized by the status quo. The time is now to build a model of shared virtual care that embraces Canada's diverse cultures and provides its older ethnic adults with access to health innovations in partnership with equally invested family caregivers who have their health at heart.

Figure 1. Medly Caregiver Live Report screenshots displaying patient partner health status and data generated on the Medly patient app.

Conflicts of Interest

QP, NE, RL, and KY are employed by the University Health Network, where the Medly and Ned applications were developed. QP owns intellectual property rights to the Ned application and is entitled to personally benefit from any commercial use of the intellectual property. AJ has no conflicts to declare.

Multimedia Appendix 1
Ned care partners tip sheet.
[PDF File (Adobe PDF File), 794 KB - jmir_v24i1e29876_app1.pdf]

References

1. Shaw J, Jamieson T, Agarwal P, Griffin B, Wong I, Bhatia RS. Virtual care policy recommendations for patient-centred primary care: findings of a consensus policy dialogue using a nominal group technique. J Telemed Telecare 2018 Oct;24(9):608-615. [doi: 10.1177/1357633X17730444] [Medline: 28945161]
2. Glazier RH, Green ME, Wu FC, Frymire E, Kopp A, Kiran T. Shifts in office and virtual primary care during the early COVID-19 pandemic in Ontario, Canada. CMAJ 2021 Feb 08;193(6):E200-E210 [FREE Full text] [doi: 10.1503/cmaj.202303] [Medline: 33558406]
3. Bhatia RS, Chu C, Pang A, Tadrous M, Stamenova V, Cram P. Virtual care use before and during the COVID-19 pandemic: a repeated cross-sectional study. CMAJ Open 2021 Feb 17;9(1):E107-E114 [FREE Full text] [doi: 10.9778/cmajo.20200311] [Medline: 33597307]
4. Demographic estimates by age and sex, provinces and territories. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/pub/71-607-x/71-607-x2020018-eng.htm#htu [accessed 2022-01-02]
5. Visible Minority (15), Generation Status (4), Age (12) and Sex (3) for the Population in Private Households of Canada, Provinces and Territories, Census Metropolitan Areas and Census Agglomerations, 2016 Census - 25% Sample Data. Statistics Canada. URL: https://www150.statcan.gc.ca/n1/en/catalogue/98-400-X2016190 [accessed 2022-01-02]
6. Immigration and Ethnocultural Diversity in Canada. Statistics Canada. URL: http://www12.statcan.gc.ca/nhs-enm/2011/as-sa/99-010-x/99-010-x2011001-eng.cfm [accessed 2022-01-02]
7. Canadian Chronic Disease Indicators (CCDI). Government of Canada. URL: https://health-infobase.canada.ca/ccdi/data-tool/?Dom=1 [accessed 2022-01-02]

8. Seniors and the Health Care System: What Is the Impact of Multiple Chronic Conditions?. Canada: Canadian Institute for Health Information; 2021 Jan 01. URL: https://secure.cihi.ca/free_products/air-chronic_disease_aib_en.pdf [accessed 2022-01-03]
9. Crawford A, Serhal E. Digital health equity and COVID-19: the innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
10. Nouri S, Khoong EC, Lyles CR, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catalyst 2020 May 04:1 [FREE Full text]
11. Immigration and ethnocultural diversity: Key results from the 2016 Census.: Statistics Canada; 2017 Oct 25. URL: https://www150.statcan.gc.ca/n1/daily-quotidien/171025/dq171025b-eng.htm?indid=14428-3&indgeo=0 [accessed 2022-01-02]
12. Proposed Standards for Race-Based and Indigenous Identity Data Collection and Health Reporting in Canada. Ottawa: Canadian Institute for Health Information; 2020. URL: https://www.cihi.ca/sites/default/files/document/proposed-standard-for-race-based-data-en.pdf [accessed 2022-01-03]
13. Defining culture.: Statistics Canada URL: https://www150.statcan.gc.ca/n1/pub/87-542-x/2011001/section/s3-eng.htm [accessed 2022-01-02]
14. López L, Green AR, Tan-McGrory A, King RS, Betancourt JR. Bridging the digital divide in health care: the role of health information technology in addressing racial and ethnic disparities. The Joint Commission Journal on Quality and Patient Safety 2011 Oct;37(10):437-445. [doi: 10.1016/s1553-7250(11)37055-9]
15. Lyles CR, Sarkar U. Health literacy, vulnerable patients, and health information technology use: where do we go from here? J Gen Intern Med 2015 Mar 15;30(3):271-272 [FREE Full text] [doi: 10.1007/s11606-014-3166-5] [Medline: 25588688]
16. Hou F, Frank K, Schimmele C. Economic impact of COVID-19 among visible minority groups. Statistics Canada. URL: https://tinyurl.com/2p98hsu3 [accessed 2022-01-03]
17. McLaren J. Racial Disparity in COVID-19 Deaths: Seeking Economic Roots with Census Data. Cambridge, MA: National Bureau of Economic Research; 2020 Jun. URL: https://www.nber.org/system/files/working_papers/w27407/w27407.pdf [accessed 2022-01-03]
18. Couch KA, Fairlie RW, Xu H. The impacts of COVID-19 on minority unemployment: first evidence from April 2020 CPS microdata. SSRN Journal 2020 May 20:1-42. [doi: 10.2139/ssrn.3604814]
19. Tai D, Shah A, Doubeni C, Sia I, Wieland M. The disproportionate impact of COVID-19 on racial and ethnic minorities in the United States. Clin Infect Dis 2021 Feb 16;72(4):703-706 [FREE Full text] [doi: 10.1093/cid/ciaa815] [Medline: 32562416]
20. Gordon NP, Hornbrook MC. Older adults' readiness to engage with eHealth patient education and self-care resources: a cross-sectional survey. BMC Health Serv Res 2018 Mar 27;18(1):220 [FREE Full text] [doi: 10.1186/s12913-018-2986-0] [Medline: 29587721]
21. Miyawaki CE. A review of ethnicity, culture, and acculturation among Asian caregivers of older adults (2000-2012). Sage Open 2015 Feb 17;5(1):215824401456636 [FREE Full text] [doi: 10.1177/2158244014566365] [Medline: 26229736]
22. Pharr JR, Dodge Francis C, Terry C, Clark MC. Culture, caregiving, and health: exploring the influence of culture on family caregiver experiences. ISRN Public Health 2014 Mar 26;2014:1-8. [doi: 10.1155/2014/689826]
23. Lai DWL. Filial piety, caregiving appraisal, and caregiving burden. Res Aging 2009 Nov 13;32(2):200-223. [doi: 10.1177/0164027509351475]
24. Lum J. In Focus: Informal Caregiving. Toronto: Ryerson University; 2011. URL: https://www.ryerson.ca/content/dam/crncc/knowledge/infocus/informalcaregiving/InFocus-InformalCaregiving.pdf [accessed 2022-01-03]
25. Hylmar S. A profile of family caregivers in Ontario. The Change Foundation 2016:1-58.
26. Berry J, Georgas J. An ecocultural perspective on cultural transmission: The family across cultures. In: Cultural Transmission: Psychological, Developmental, Social, and Methodological Aspects. Cambridge, MA: Cambridge University Press; Jun 05, 2012.
27. Jones PS, Winslow BW, Lee JW, Burns M, Zhang XE. Development of a caregiver empowerment model to promote positive outcomes. J Fam Nurs 2011 Feb 22;17(1):11-28. [doi: 10.1177/1074840710394854] [Medline: 21343620]
28. Penrod J, Baney B, Loeb S, McGhan G, Shipley P. The influence of the culture of care on informal caregivers' experiences. ANS Adv Nurs Sci 2012;35:76. [doi: 10.1097/ans.0b013e318244555a]
29. Cheng S, Chan ACM. Filial piety and psychological well-being in well older Chinese. J Gerontol B Psychol Sci Soc Sci 2006 Sep 01;61(5):P262-P269. [doi: 10.1093/geronb/61.5.p262] [Medline: 16960229]
30. Wang J, Liang H, Wu B. Association between filial piety and caregiving burden among Chinese adults' children caregivers. Innovation in Aging 2017;1(suppl_1):1279. [doi: 10.1093/geroni/igx004.4667]
31. Hsueh K, Bachman JA, Richardson LI, Cheng W, Zimmerman RS. Mechanisms through which reciprocal filial values protect wellbeing of Chinese adult-child caregivers in the US. Res Nurs Health 2014 Apr 14;37(2):155-166. [doi: 10.1002/nur.21585] [Medline: 24420507]
32. Visible Minority (15), Highest Certificate, Diploma or Degree (15), Generation Status (4), Age (9) and Sex (3) for the Population Aged 15 Years and Over in Private Households of Canada, Provinces and Territories and Census Metropolitan

Areas, 2016 Census - 25% Sample Data. Data tables, 2016 Census.: Statistics Canada; 2016. URL: https://tinyurl.com/3pz4nks8 [accessed 2022-01-02]
33. Rootman I, Gordon-El-Bihbety D. A Vision for a Health Literate Canada: Report of the Expert Panel on Health Literacy. Canadian Public Health Association. 2008. URL: https://swselfmanagement.ca/uploads/ResourceDocuments/CPHA%20(2008)%20A%20Vision%20for%20a%20Health%20Literate%20Canada.pdf [accessed 2022-01-03]
34. Huynh A, Malli N. Levelling up: the quest for digital literacy. Canada: Brookfield Institute for Innovation & Entrepreneurship; 2018. URL: https://www.voced.edu.au/content/ngv:80067 [accessed 2022-01-03]
35. Intergenerational Education Mobility and Labour Market Outcomes: Variation Among the Second Generation of Immigrants in Canada.: Statistics Canada URL: https://www150.statcan.gc.ca/n1/pub/11f0019m/11f0019m2019006-eng.htm [accessed 2021-03-12]
36. Adelman RD, Tmanova LL, Delgado D, Dion S, Lachs MS. Caregiver burden: a clinical review. JAMA 2014 Mar 12;311(10):1052-1060. [doi: 10.1001/jama.2014.304] [Medline: 24618967]
37. Gérain P, Zech E. Informal caregiver burnout? Development of a theoretical framework to understand the impact of caregiving. Front Psychol 2019 Jul 31;10:1748 [FREE Full text] [doi: 10.3389/fpsyg.2019.01748] [Medline: 31428015]
38. Bastawrous M. Caregiver burden--a critical discussion. Int J Nurs Stud 2013 Mar;50(3):431-441. [doi: 10.1016/j.ijnurstu.2012.10.005] [Medline: 23131724]
39. National Academies of Sciences, Engineering, and Medicine. Economic Impact of Family Caregiving. In: Families Caring for an Aging America. Washington, DC: National Academies Press (US); 2016.
40. Razani J, Kakos B, Orieta-Barbalace C, Wong J, Casas R, Lu P, et al. Predicting caregiver burden from daily functional abilities of patients with mild dementia. J Am Geriatr Soc 2007 Sep;55(9):1415-1420 [FREE Full text] [doi: 10.1111/j.1532-5415.2007.01307.x] [Medline: 17767684]
41. Conway K. The experience of adult children caregiving for aging parents. Home Health Care Management & Practice 2018 Oct 08;31(2):92-98. [doi: 10.1177/1084822318803559]
42. Bastawrous M, Gignac MA, Kapral MK, Cameron JI. Factors that contribute to adult children caregivers' well-being: a scoping review. Health Soc Care Community 2015 Sep 04;23(5):449-466. [doi: 10.1111/hsc.12144] [Medline: 25472851]
43. Cook SK, Snellings L, Cohen SA. Socioeconomic and demographic factors modify observed relationship between caregiving intensity and three dimensions of quality of life in informal adult children caregivers. Health Qual Life Outcomes 2018 Aug 29;16(1):169 [FREE Full text] [doi: 10.1186/s12955-018-0996-6] [Medline: 30157852]
44. Lai DWL, Luk PKF, Andruske CL. Gender differences in caregiving: a case in Chinese Canadian caregivers. Journal of Women & Aging 2008 Oct 17;19(3-4):161-178. [doi: 10.1300/j074v19n03_11]
45. Mitnick S, Leffler C, Hood VL, American College of Physicians Ethics, Professionalism and Human Rights Committee. Family caregivers, patients and physicians: ethical guidance to optimize relationships. J Gen Intern Med 2010 Mar 9;25(3):255-260 [FREE Full text] [doi: 10.1007/s11606-009-1206-3] [Medline: 20063128]
46. National Academies of Sciences, Engineering, and Medicine. Family Caregivers' Interactions with Health Care and Long-Term Services and Supports. In: Families Caring for an Aging America. Washington, DC: National Academies Press (US); 2016.
47. Latulipe C, Quandt SA, Melius KA, Bertoni A, Miller DP, Smith D, et al. Insights into older adult patient concerns around the caregiver proxy portal use: qualitative interview study. J Med Internet Res 2018 Nov 02;20(11):e10524 [FREE Full text] [doi: 10.2196/10524] [Medline: 30389654]
48. Casillas A, Cemballi AG, Abhat A, Lemberg M, Portz JD, Sadasivaiah S, et al. An untapped potential in primary care: semi-structured interviews with clinicians on how patient portals will work for caregivers in the safety net. J Med Internet Res 2020 Jul 20;22(7):e18466 [FREE Full text] [doi: 10.2196/18466] [Medline: 32706709]
49. Piette JD, Valverde H, Marinec N, Jantz R, Kamis K, de la Vega CL, et al. Establishing an independent mobile health program for chronic disease self-management support in bolivia. Front Public Health 2014 Aug 13;2:95 [FREE Full text] [doi: 10.3389/fpubh.2014.00095] [Medline: 25165687]
50. Silveira MJ, Given CW, Cease KB, Sikorskii A, Given B, Northouse LL, et al. Cancer Carepartners: Improving patients' symptom management by engaging informal caregivers. BMC Palliat Care 2011 Nov 25;10(1):1-11. [doi: 10.1186/1472-684x-10-21]
51. Bouldin ED, Aikens JE, Piette JD, Trivedi RB. Relationship and communication characteristics associated with agreement between heart failure patients and their Carepartners on patient depressive symptoms. Aging Ment Health 2019 Sep 20;23(9):1122-1129 [FREE Full text] [doi: 10.1080/13607863.2018.1481923] [Medline: 30569750]
52. Piette J, Striplin D, Marinec N, Chen J, Aikens J. A randomized trial of mobile health support for heart failure patients and their informal caregivers: impacts on caregiver-reported outcomes. Medical Care 2015;53:699. [doi: 10.1097/mlr.0000000000000378]
53. Piette J, Striplin D, Aikens J, Lee A, Marinec N, Mansabdar M, et al. Impacts of post-hospitalization accessible health technology and caregiver support on 90-day acute care use and self-care assistance: a randomized clinical trial. Am J Med Qual 2021;36(3):145-155. [doi: 10.1177/1062860620943673] [Medline: 32723072]

54. Piette JD, Striplin D, Marinec N, Chen J, Trivedi RB, Aron DC, et al. A mobile health intervention supporting heart failure patients and their informal caregivers: a randomized comparative effectiveness trial. J Med Internet Res 2015 Jun 10;17(6):e142 [FREE Full text] [doi: 10.2196/jmir.4550] [Medline: 26063161]
55. Piette JD, Marinec N, Janda K, Morgan E, Schantz K, Yujra ACA, et al. Structured caregiver feedback enhances engagement and impact of mobile health support: a randomized trial in a lower-middle-income country. Telemed J E Health 2016 Apr;22(4):261-268 [FREE Full text] [doi: 10.1089/tmj.2015.0099] [Medline: 26352854]
56. eHealth Innovation. URL: https://ehealthinnovation.org/ [accessed 2021-04-22]
57. Medly. URL: https://medly.ca/ [accessed 2021-04-22]
58. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Developing healthcare rule-based expert systems: case study of a heart failure telemonitoring system. Int J Med Inform 2012 Aug;81(8):556-565. [doi: 10.1016/j.ijmedinf.2012.03.001] [Medline: 22465288]
59. Ware P, Ross HJ, Cafazzo JA, Boodoo C, Munnery M, Seto E. Outcomes of a heart failure telemonitoring program implemented as the standard of care in an outpatient heart function clinic: pretest-posttest pragmatic study. J Med Internet Res 2020 Feb 08;22(2):e16538 [FREE Full text] [doi: 10.2196/16538] [Medline: 32027309]
60. Ware P, Dorai M, Ross HJ, Cafazzo JA, Laporte A, Boodoo C, et al. Patient adherence to a mobile phone-based heart failure telemonitoring program: a longitudinal mixed-methods study. JMIR mHealth uHealth 2019 Feb 26;7(2):e13259 [FREE Full text] [doi: 10.2196/13259] [Medline: 30806625]
61. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. Evaluating the implementation of a mobile phone-based telemonitoring program: longitudinal study guided by the consolidated framework for implementation research. JMIR mHealth uHealth 2018 Jul 31;6(7):e10768 [FREE Full text] [doi: 10.2196/10768] [Medline: 30064970]
62. Ned Clinic. URL: https://nedclinic.ca/ [accessed 2021-04-22]
63. Pham Q, Cafazzo JA, Feifer A. Adoption, acceptability, and effectiveness of a mobile health app for personalized prostate cancer survivorship care: protocol for a realist case study of the ned app. JMIR Res Protoc 2017 Oct 12;6(10):e197 [FREE Full text] [doi: 10.2196/resprot.8051] [Medline: 29025699]
64. Pham Q, Hearn J, Bender JL, Berlin A, Brown I, Bryant-Lukosius D, et al. Virtual care for prostate cancer survivorship: protocol for an evaluation of a nurse-led algorithm-enhanced virtual clinic implemented at five cancer centres across Canada. BMJ Open 2021 Apr 21;11(4):e045806 [FREE Full text] [doi: 10.1136/bmjopen-2020-045806] [Medline: 33883153]
65. Ethnic diversity and immigration.: Statistics Canada URL: https://www150.statcan.gc.ca/n1/pub/11-402-x/2011000/chap/imm/imm-eng.htm [accessed 2021-02-25]

Abbreviations

IVR: interactive voice response

Edited by A Mavragani; submitted 23.04.21; peer-reviewed by N Gordon, M Wasilewski, E Ray Chaudhuri, W Tang; comments to author 08.10.21; revised version received 18.10.21; accepted 01.12.21; published 07.01.22.

Please cite as:
Pham Q, El-Dassouki N, Lohani R, Jebanesan A, Young K
The Future of Virtual Care for Older Ethnic Adults Beyond the COVID-19 Pandemic
J Med Internet Res 2022;24(1):e29876
URL: https://www.jmir.org/2022/1/e29876
doi:10.2196/29876
PMID:34994707

Viewpoint

Applying Human-Centered Design Principles to Digital Syndromic Surveillance at a Mass Gathering in India: Viewpoint

Ahmed Shaikh[1*], MD; Abhishek Bhatia[2,3*], MS; Ghanshyam Yadav[4*], MD; Shashwat Hora[5*], MBBS, MBID; Chung Won[6*], MD; Mark Shankar[7*], MD, MBA; Aaron Heerboth[3*], MD; Prakash Vemulapalli[8*], MD, MBA; Paresh Navalkar[9*], MD, MBA; Kunal Oswal[10*], BDS, MPH; Clay Heaton[3*], BA, MS, SM; Sujata Saunik[11,12*], MA; Tarun Khanna[13*], PhD; Satchit Balsari[3,14,15*], MD, MPH

[1]Institute for Critical Care Medicine, Mount Sinai Hospital, New York, NY, United States
[2]Carolina Health Informatics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States
[3]India Digital Health Network, Lakshmi Mittal and Family South Asia Institute, Harvard University, Cambridge, MA, United States
[4]Department of Obstetrics and Gynecology, Baylor College of Medicine, Houston, TX, United States
[5]Articulate Labs, Inc, San Francisco, CA, United States
[6]Department of Emergency Medicine, Memorial Hermann Hospital -Baylor College of Medicine, Houston, TX, United States
[7]Department of Emergency Medicine, Jacobi Medical Center, New York, NY, United States
[8]University Hospitals Center for Emergency Medicine, Cleveland Medical Center, Cleveland, OH, United States
[9]Lifesupporters Institute of Health Sciences, Mumbai, India
[10]Department of Public Health Dentistry, Sharad Pawar Dental College, Maharashtra, India
[11]Department of General Administration, Government of Maharashtra, Mumbai, India
[12]Harvard TH Chan School of Public Health, Boston, MA, United States
[13]Harvard Business School, Boston, MA, United States
[14]Department of Global Health and Population, Harvard TH Chan School of Public Health, Boston, MA, United States
[15]Department of Emergency Medicine, Beth Israel Deaconess- Harvard Medical School, Boston, MA, United States
[*]all authors contributed equally

Corresponding Author:
Satchit Balsari, MD, MPH
Department of Global Health and Population
Harvard TH Chan School of Public Health
651 Huntington Avenue
703C
Boston, MA, 02115
United States
Phone: 1 6174951000
Email: balsari@hsph.harvard.edu

Abstract

In the wake of the COVID-19 pandemic, digital health tools have been deployed by governments around the world to advance clinical and population health objectives. Few interventions have been successful or have achieved sustainability or scale. In India, government agencies are proposing sweeping changes to India's digital health architecture. Underpinning these initiatives is the assumption that mobile health solutions will find near universal acceptance and uptake, though the observed reticence of clinicians to use electronic health records suggests otherwise. In this practice article, we describe our experience with implementing a digital surveillance tool at a large mass gathering, attended by nearly 30 million people. Deployed with limited resources and in a dynamic chaotic setting, the adherence to human-centered design principles resulted in near universal adoption and high end-user satisfaction. Through this use case, we share generalizable lessons in the importance of contextual relevance, stakeholder participation, customizability, and rapid iteration, while designing digital health tools for individuals or populations.

(J Med Internet Res 2022;24(1):e27952) doi:10.2196/27952

KEYWORDS

mHealth; design; human centered design; intervention; syndromic surveillance; digital health

Introduction

In 2020, the government of India announced the National Digital Health Mission (NDHM), a vision for a federated health information ecosystem that is expected to catalyze India's digital health transformation [1]. The NDHM distinguishes itself from its predecessors by embracing an enabling platform approach that is expected to allow market solutions like mobile apps and wearables to seamlessly interface with patient health records, opening the possibility of exponential growth in digital solutions in both the public and private sectors [2-4].

Though the proposed open ecosystem provides an unprecedented opportunity for growth and competition in the development of digital health tools, the lack of systematic approaches to evaluate and validate digital health interventions risks diluting the impact of the very large investments expected in this space [5-7]. The vast majority of digital health interventions fail to scale for a variety of reasons, including the lack of human-centered design and poor contextual knowledge [8,9]. These challenges are further exacerbated in low resource settings that suffer from significant institutional voids [10,11]. The near-universal spectacle of overburdened providers entering data in hard-to-navigate survey forms points to the emphasis placed on programmatic and reporting needs, instead of the needs of the key users, namely, patients and health care providers [12,13].

Solutions that have succeeded have demonstrated a commitment to human-centered design and a deep understanding of the infrastructural, social, and economic constraints faced by providers in primary care settings. Programs like Gujarat's TECHO platform for maternal and child health services in tribal communities and Sangath's ESSENCE program for training community health workers in mental health have relied on a resource-intensive strategy that combines task-shifting, training, and technology for successful digital health implementation in India [14,15]. These programs adopted iterative ideation, prototyping, and testing cycles early in their product design process [16,17].

In this paper, we describe the successful implementation of a mobile health tool in a dynamic transient setting, the 2015 Nashik Kumbh Mela in India, where the administrators had minimal to no time allocated to training providers, and yet, adoption was near universal. The tool was used by over 100 clinicians to enter data and track performance, and by state administrators to monitor disease outbreaks in real-time [18]. The tool also provided essential data on utilization, resource-allocation, and prescription patterns. We attribute the tool's success to adherence to human-centered design principles, great emphasis on user experience, meticulous attention to workflow, and strict adherence to data minimization, principles which may serve well the many large-scale digital health rollouts being attempted all over the developing world [19-22]. Through this use case, we share generalizable lessons in the importance of contextual relevance, stakeholder participation, customizability, and rapid iteration, while designing digital health tools for individuals or populations.

The Nashik Kumbh Mela

The Kumbh Mela is a religious mass gathering that occurs every 3 to 4 years at 1 of 4 pilgrimage sites in India [23]. In 2015, the Kumbh Mela was held in the adjacent towns of Nashik (population: 1,486,053) and Trimbakeshwar (population: 12,056) from August 26 to September 25, 2015, and attended by an estimated 30 million people [24,25]. A span of three 3-day festivities at each site marked the most auspicious days when the majority of pilgrims seek a dip in the holy waters of the Ram Kund (pond) in Nashik or the Godavari river in Trimbakeshwar.

As with all Kumbh Melas, the state government serves as patron and host, providing a range of services to ensure the well-being and safety of the visitors and the local population. Preparations begin months in advance and entail additional transport facilities; staging of incoming traffic away from crowded zones; provision of clean water supply, electricity, and temporary shelter; and construction of over 50 temporary health care facilities along the arterial routes leading to the pilgrimage sites [26,27]. These clinics are staffed by a physician, nurse, and pharmacist, and offer little to no laboratory tests. The concern for stampedes and disease outbreaks is high at the Kumbh Melas, and governments have historically invested significant resources toward mitigating the risks for both [28-32]. Strict water quality monitoring, provision of clean safe drinking water, and provision of health outposts (clinics) have been the norm at these Melas [30]. Physicians posted at these clinics routinely have no more than a couple of minutes to see each patient, and hastily scribble down the patient's chief complaint on a paper-based log.

In 2015, the Public Health Department, Government of Maharashtra, invited our team of researchers to build on prior experience at the Allahabad Mela, to create a digital disease surveillance system for the 2015 Kumbh Mela [33].

Designing for Digital Health

We adopted design principles recommended by IDEO's Field Guide to Human-Centered Design [34]. These principles also appear in the more recently published "Human-Centered Design 4 Health" guidelines by UNICEF and in the World Health Organization's "Digital Health Implementation Investment Guide" [21,22,35], all of which underscore the importance of an approach and recognize that projects loop through the phases of ideation, inspiration, and implementation multiple times as they incorporate user feedback in every stage until a refined solution is ready for scale [17].

Step 1: Inspiration

Assembling an Interdisciplinary Team

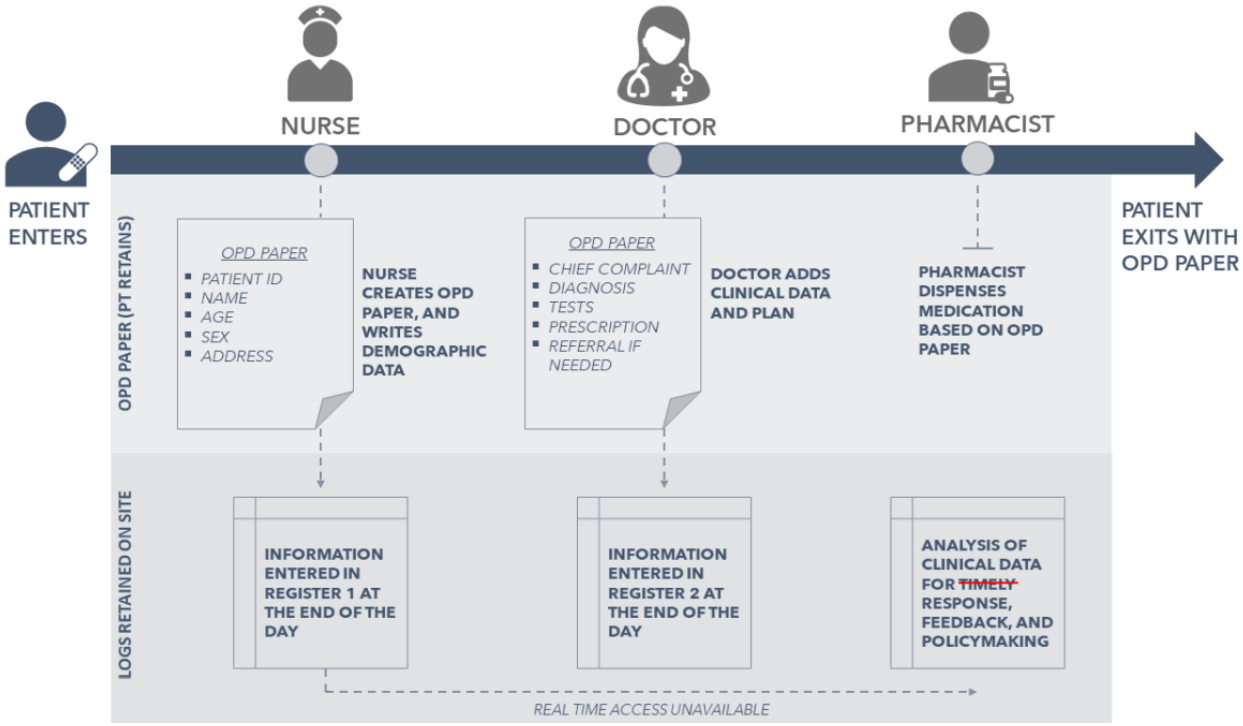
The process of inspiration entails an exploration of the circumstances that motivate the need for a solution [17]. At the start of the design process, we defined the following 2 groups of end users: (1) the clinicians staffing the clinics, who would also be responsible for data entry, and (2) the public health officials charged with managing potential disease outbreaks. We assembled an interdisciplinary team that included senior

state officials (to facilitate subsequent approvals) and subject-matter experts, including field researchers, data scientists, clinicians, and medical students with prior experience at the 2013 Kumbh. This team composition ensured the technical team had access to rich contextual intelligence from the early stages of the design process and to senior policymakers.

Needs Assessment

The research team then carried out a comprehensive review of prior research and related literature; conducted a series of key informant interviews with local and state public health leadership and medical officers; and performed site visits to understand the layout of the clinics, workflow, potential pitfalls, and day-to-day logistics (Figure 1).

Figure 1. Workflow at the 2015 Kumbh Mela clinics prior to the intervention. OPD: outpatient department.



During the Melas, each patient encounter is typically recorded on a sheet of paper known as the "OPD" (outpatient department) paper, a near ubiquitous document in all state-run primary care clinics in India. The form used in Maharashtra has not changed since 1933 and is typically completed for the patient by the nurse and carried by the patient first to the doctor and then to the pharmacist, who retrieves the document while dispensing medication. Various parameters, such as medications dispensed and clinical diagnoses, are tallied manually and entered into a log by the nurses and pharmacists at the end of the workday. The log is then physically transported to a nodal office where similar reports from all the temporary clinics are totaled and entered into a spreadsheet that is then transmitted to other administrators. On busy days, the staff spend hours (over time) tallying the spreadsheets late in the night to provide actionable analysis the next morning.

The inefficiency and possible inaccuracy associated with the system prompted India's National Disaster Management Authority to pilot a tablet-based system at the Allahabad Mela in 2013 [18].In 2015, the then principal secretary of public health in Maharashtra asked for the disease surveillance system to be digitized. Acknowledging the significant human labor associated with tallying the data manually every day, the local administrative team in Nashik was strongly in favor of digitizing the data, but skeptical about the participation of clinical staff. The clinicians were cautious and guarded about embracing a new digital tool, fearing that it would make their job harder.

Step 2: Ideation

Common Themes

Affinity diagrams resulting from open-ended discussions with multiple stakeholders helped identify the following key themes: the tool would have to offer more than simply a digital medium to collect data; it would necessarily have to make the jobs of the clinical staff easier, not harder; it would have to save time; it would need to avoid redundancy; it would need to result in actionable, reliable, verifiable, and timely analysis; the clinicians entering the data would need to see the benefit; it would need to have little to no barriers for onboarding users; and adoption and retention would need to depend on the clinician's user experience [36]. These observations led to the design decisions described below.

Data Minimization

The existing paper-based system relegated clinical and ancillary staff to duplicating a low-yield, inefficient, clerical task. In the absence of confirmatory tests, syndromic surveillance would only require daily incidence of presenting complaints. Given the dynamic population flux in and out of the city, a spike in absolute numbers could merely reflect a transient rise in the visiting population. For effective surveillance, the relative risk of one disease compared to the incidence of others would be

helpful. The only data points that were strictly necessary for such syndromic surveillance were comprehensive tallies of all presenting complaints. Age, gender, and the location of the clinic would also be informative. Other information like patients' social histories, the treatments they were given, and even their vital signs, while all essential for documenting a good clinical encounter, are unnecessary for syndromic surveillance, especially in a resource-constrained environment with little analytic or response capacity to use the additional information.

Prototyping

To find a solution that balanced fidelity, speed, and cost, this stage of the ideation process involved co-creation to ensure that end-user needs informed design choices (Figure 2) [16,37].

Figure 2. Understanding the layout, context, and disease surveillance needs from public health officials at the start of the Mela.

Historically, having only used tallied data from spreadsheets, local officials struggled to envision dashboard designs and requested a printable version of the "table" they were most familiar with. In order to overcome design fixation, we adopted parallel prototyping to develop the following 2 sets of outputs: a data-entry tablet-based interface and a visualization tool for public health officials.

Tablet Interface

To optimize the tool's user interface and user experience, clinicians who would staff the Mela clinics were involved in co-creating the data-entry tool. The design team met with over 100 doctors and nurses during their preparatory training sessions a month prior to the Mela. Following an orientation section that described the intervention at the 2013 Mela and the potential for real-time analytics for census, inventory, and disease surveillance, the clinical staff were offered usability testing in small groups. Users were invited to provide feedback on the perceived utility and usability of the tool, with particular attention to the interface (positioning; font type and size; and entry choices including checkboxes, radio buttons, steppers, toggle switches, dropdowns, and list boxes). This process informed iterative cycles of creation and testing. The inclusion of end users in this dynamic process helped to foster their sense of ownership in the process and ultimately resulted in high adoption rates. The final tool collected only 4 data points necessary for surveillance and clinical care (the medical record number [MRN], age, sex, and chief complaint). The MRN was structured such that it revealed clinic location and date, both of which were autopopulated in the digital tool, requiring the physician to only enter the last 3 digits, and this greatly minimized data-entry errors (Figure 3). A total of 42 presenting complaints were listed in a drop-down menu, in order of expected frequency, with autocompletion text options and with the buttons, font size, and positioning of modules tailored to physician needs. Date, time, and GPS data were autopopulated from the tablet data. This degree of attention to the user experience was necessary to reduce the number of clicks required of physicians and to reduce their cognitive burden [20]. The abbreviated OPD paper contained a field for the registration number, and a checklist of diagnoses and the most commonly prescribed medications (Figure 4).

Figure 3. An initial mockup of the proposed EMcounter tool, minimizing data entry requirements for providers at the Mela while reducing errors in reporting for real-time epidemiological surveillance.

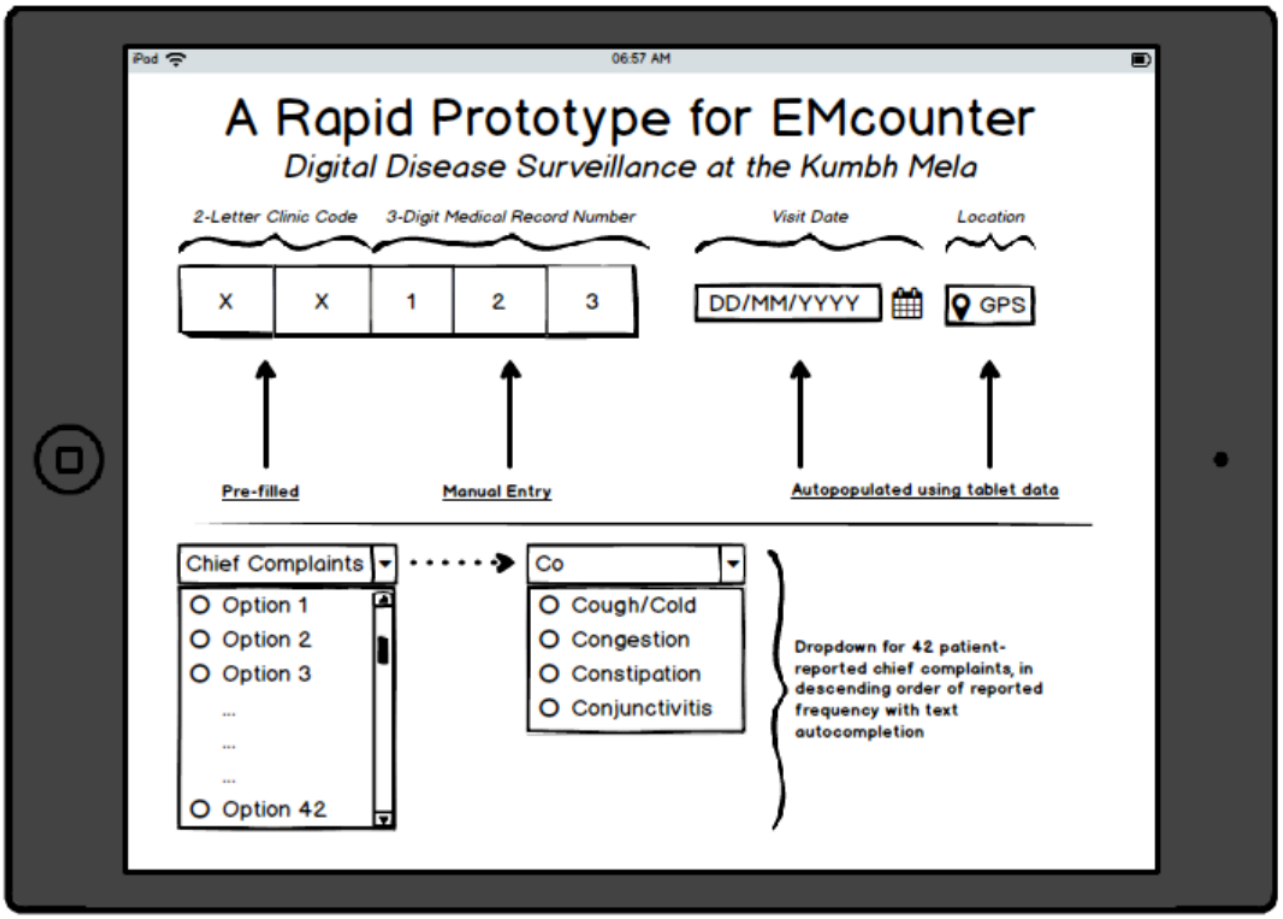
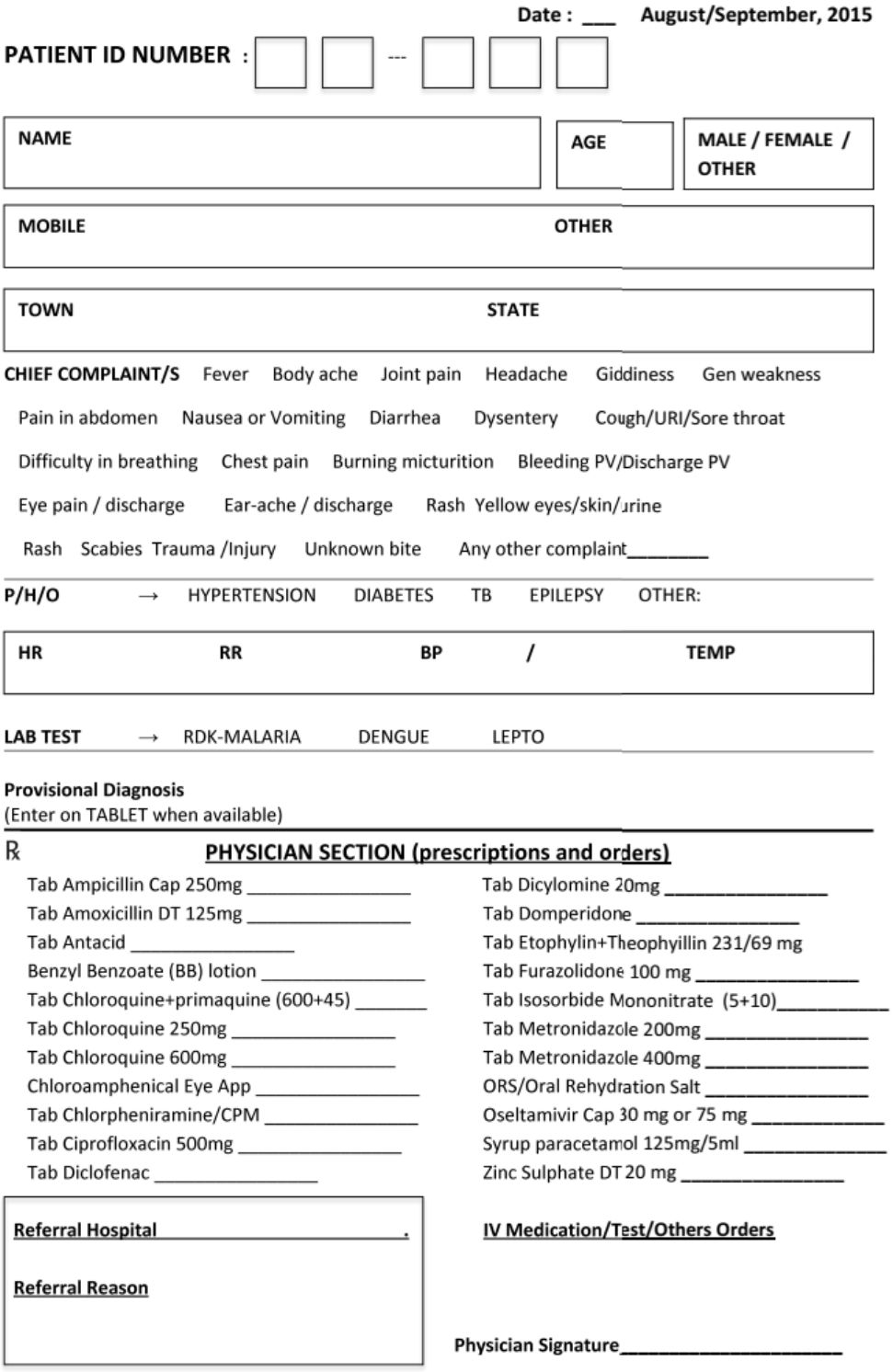


Figure 4. The OPD paper with structured sections and tailored response options, abbreviated from the free-text OPD sheet provided preintervention to providers at the Mela. OPD: outpatient department.


Date : ___ August/September, 2015

PATIENT ID NUMBER : ☐ ☐ --- ☐ ☐ ☐

NAME | AGE | MALE / FEMALE / OTHER

MOBILE OTHER

TOWN STATE

CHIEF COMPLAINT/S Fever Body ache Joint pain Headache Giddiness Gen weakness

Pain in abdomen Nausea or Vomiting Diarrhea Dysentery Cough/URI/Sore throat

Difficulty in breathing Chest pain Burning micturition Bleeding PV/Discharge PV

Eye pain / discharge Ear-ache / discharge Rash Yellow eyes/skin/urine

Rash Scabies Trauma /Injury Unknown bite Any other complaint_____

P/H/O → HYPERTENSION DIABETES TB EPILEPSY OTHER:

HR RR BP / TEMP

LAB TEST → RDK-MALARIA DENGUE LEPTO

Provisional Diagnosis
(Enter on TABLET when available)

℞ PHYSICIAN SECTION (prescriptions and orders)

Tab Ampicillin Cap 250mg _____
Tab Amoxicillin DT 125mg _____
Tab Antacid _____
Benzyl Benzoate (BB) lotion _____
Tab Chloroquine+primaquine (600+45) _____
Tab Chloroquine 250mg _____
Tab Chloroquine 600mg _____
Chloroamphenical Eye App _____
Tab Chlorpheniramine/CPM _____
Tab Ciprofloxacin 500mg _____
Tab Diclofenac _____

Tab Dicylomine 20mg _____
Tab Domperidone _____
Tab Etophylin+Theophyillin 231/69 mg
Tab Furazolidone 100 mg _____
Tab Isosorbide Mononitrate (5+10)_____
Tab Metronidazole 200mg _____
Tab Metronidazole 400mg _____
ORS/Oral Rehydration Salt _____
Oseltamivir Cap 30 mg or 75 mg _____
Syrup paracetamol 125mg/5ml _____
Zinc Sulphate DT 20 mg _____

Referral Hospital _____.

Referral Reason

IV Medication/Test/Others Orders

Physician Signature_____

Dashboards

In addition to the tables requested by the public health officials, the final product included downloadable tables (Multimedia Appendix 1), as well as a real-time interactive dashboard that allowed the user to query and filter the data via several permutations of location, customized timeframe, age group, gender, and chief complaint (Multimedia Appendix 1). This would allow real-time epidemiological exploration of any observed atypical trends, triggering the public health system to rapidly launch an inquiry or response.

Field Testing

The 50 temporary clinics across Nashik and Trimbakeshwar included temporary structures made from cloth, bamboo, and rope; repurposed rooms in existing buildings; and even designated areas in cavernous temple halls buttressed by thick stone walls, impermeable to cell phone signals (Figure 5). The system was designed to transmit data over 3G and 4G

connections via SIM-enabled tablet computers to the cloud-based analytic tool, remotely accessible with authentication. The field visits revealed that several sites did not have good cell phone coverage, requiring that the devices be periodically swapped and brought to an area with good signal. The clinics had few power outlets, and they were often away from the doctor's desk, necessitating that each tablet computer be accompanied by an extra battery pack and a multipoint extension cord.

Figure 5. A pre-Mela site visit revealing that some clinics would be held in spaces with thick stone walls, precluding cellular service for real-time reporting of collected data.

Step 3: Implementation

Preparation

The co-creation processes helped the design team preempt and address issues that normally account for poor user compliance, including poor perceived utility, lack of incentive, and suboptimal user experience. The ideation process and field visits allowed us to preempt a series of logistical and infrastructural issues, resulting in organized "deployment kits" for each clinic site. The kit included a battery pack, extension power cord, tablet computer, user manual, sign-out sheet, and phone number to a 24/7 helpline of medical student volunteers trained in troubleshooting the software (Figure 6).

Figure 6. The EMCounter kit provided to each clinical team with labeled tablets, power supply, instructions for troubleshooting, and contact information of designated support team members.

Training and Support

A 1-hour training session was allocated by the public health department for demonstrating and using the tool, on the eve of the Mela. Most of the physicians had seen earlier iterations during the co-creation sessions and were able to test the tool with little to no supervision (Figure 7). Many of the pharmacists and nurses did not receive any training, were unaware of the project, and were trained on-site the next morning.

Figure 7. Core team member Dr John Won introduces the 2015 digital surveillance effort.

The troubleshooting team comprised voluntary medical students (from India) and medical residents (from New York), all of whom had co-created the tool, and several of whom had worked at the 2013 Kumbh Mela and could anticipate the uncertainties associated with deployment in a large chaotic mass gathering. The government endorsed the software as the "official" data collection tool and circulated an official state memo to all providers mandating that they adopt the tool. Pairs of volunteers visited every site, twice daily, and checked in with every provider via phone (Figure 8). WhatsApp groups were used to send out checklists for every shift, and solicit daily and exit feedback. A core team stationed at a nearby hotel with a good Wi-Fi–based internet connection was in constant communication with the technical team in Boston and Mumbai, and with public health officials in Nashik and Mumbai, providing frequent updates and helping public health officials interpret and navigate the dashboard if necessary (Figure 9). During the site visit, clinicians were provided time to explore the dashboard to examine disease epidemiology, clinic population demography, and their own census compared to their peers.

Figure 8. A volunteer medical student from the troubleshooting team visits a tent clinic at the Mela as part of daily in-person check-ins to maintain data quality and debug any issues with the tool.

Figure 9. Core team members Dr Ahmed Shaikh and Dr Shashwat Hora interpret the real-time epidemiological data presented on dashboards, along with public health officials in Nashik.

Deployment

The EMcounter digital surveillance system was deployed at over 40 clinics across Nashik and Trimbakeshwar, and recorded 33,305 discrete patient encounters over 9 days. Compliance reached 100% by the end of the first day. Local public health officials learned to query the online dashboard and routinely consulted the core team for clarifications. On day 3, noticing a spike in diarrheal diseases at a particular clinic, public health officials dispatched a team of sanitation engineers to test water at all surrounding taps for contamination. A candid testimonial about the incident from one of the clinicians is available in Multimedia Appendix 2.

Mid-project feedback revealed that 100% of the users found value in the exercise and the analytics, and the ancillary staff stopped tallying paper records after initially comparing manually maintained logs to the reported data. The census data were particularly useful in addressing supply-demand mismatches, as not all clinics saw equal footfall despite being initially staffed uniformly

Discussion

Digital disease surveillance has now been deployed by our team at 2 of the world's largest mass gatherings, under conditions of extreme uncertainty and chaos. We attribute the tool's success

in 2015 to strict adherence to design principle standards. The interdisciplinary nature of our team that included public health practitioners, physicians, computer scientists, an embedded journalist, a filmmaker, a senior bureaucrat, and several end users, precluded early design fixation and allowed us to draw upon our expertise and the carefully implemented ideation process to generate a variety of prototypes for the final users to choose from. Co-creation also instilled a sense of buy-in from an otherwise underinvested group of clinicians who were redeployed for this job from across the state. Most importantly, taking the time to explain the epidemiological rationale for the intervention encouraged buy-in. This is seldom done with digital health programs that are rolled out at scale in either the public or private sector. The lack of co-creation and the absence of user buy-in and ownership result in poor compliance and rapid attrition, incorrectly leading to the conclusion that health care providers are resistant to change [37,38]. Data minimization will also become increasingly important as recent advances in India's digital health ecosystem are likely to spur the collection of vast amounts of data. Data minimization, in addition to improving the provider experience, is also a sound privacy-preserving strategy [39,40].

This project had 2 significant limitations that may have further precluded large-scale adoption of the tool. There was no mandate to adopt any interoperability standards, as the project, despite being government sanctioned, was perceived by some as an external academic intervention. Since there were no existing integrated medical records in the public sector, this limitation was less consequential. Despite its significant success in meeting its articulated objectives, this project suffered from "pilotitis," a common fate of digital health interventions everywhere [37]. The intervention was not expanded to all primary care sites in the state as was originally envisioned. A lack of institutional memory is particularly heightened in India due to the rapid turnover of officials. Table 1 compares the project's performance against standards and principles for successful digital health implementation recommended by the recently published World Health Organization's Digital Implementation Investment Guide [21].

Table 1. Comparison of EMcounter at the Kumbh Mela, implemented in 2015, with the World Health Organization's Digital Implementation Investment Guide Checklist, released in 2020.

World Health Organization's Digital Implementation Investment Guide	Correlation	Comment
Design with the user	High	Co-created with end users
Understand the existing ecosystem	High	Embedded team members
Design for scale	High	Light back-end and low-footprint technology used
Build for sustainability	Low	Poor buy-in from some stakeholders; lack of interoperability standards would hamper integration
Be data driven	High	Statistically sound analytics
Use open standards, open data, open source, and open innovation	Medium	Interoperability standards not adopted
Reuse and improve	High	Built on innovations in a prior Mela
Address privacy and security	High	Data anonymized at source
Be collaborative	High	Multidisciplinary international team

Conclusion

Digital interventions fail when they ignore the complexity of health care interventions [36]. Unlike other sectors, there is wide variation in clinical practice from provider to provider, an even greater variation in workflow and routine among health care sites, and vast differences in health-seeking behaviors among patients, which are influenced by socioeconomic conditions, gender, age, and health literacy [41-43]. It is therefore imperative that investments in digital health projects underscore the need to co-create and pilot interventions before sanctioning their use at scale. Sandboxing, monitoring, and impact evaluation should be integral components of early design [44].

The plethora of digital applications, especially digital contact tracing solutions, adopted and deployed during the pandemic by governments around the world, despite little evidence that they work, highlights the importance of a deliberate thoughtful approach to deploying such interventions. There is little reason to not hold digital tools to the same high standard of scientific rigor as any other public health intervention.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The EMcounter dashboard.
[MP4 File (MP4 Video), 9653 KB - jmir_v24i1e27952_app1.mp4]

Multimedia Appendix 2
Clinician testimonial about the use of EMcounter for syndromic surveillance at the Kumbh Mela.
[MP4 File (MP4 Video), 19129 KB - jmir_v24i1e27952_app2.mp4]

References

1. Ayushman Bharat Digital Mission. National Health Portal. URL: https://www.nhp.gov.in/national-digital-health-mission-(ndhm)_pg [accessed 2021-05-20]
2. Ayushman Bharat. KPMG. 2019 Dec. URL: https://assets.kpmg/content/dam/kpmg/in/pdf/2019/12/universal-health-coverage-ayushman-bharat.pdf [accessed 2021-05-20]
3. 10 million screened for common diseases under Ayushman Bharat scheme. Hindustan Times. 2019. URL: https://www.hindustantimes.com/india-news/10-million-screened-for-common-diseases-under-ayushman-bharat-scheme/story-jRilUYWeolc7JQf9yCPLfL.html [accessed 2020-11-01]
4. Tuba M, Akashe S, Joshi A. ICT Systems and Sustainability. Singapore: Springer; 2020.
5. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med 2013;10(2):e1001382 [FREE Full text] [doi: 10.1371/journal.pmed.1001382] [Medline: 23424286]
6. Agarwal S, Lefevre AE, Labrique AB. A Call to Digital Health Practitioners: New Guidelines Can Help Improve the Quality of Digital Health Evidence. JMIR Mhealth Uhealth 2017 Oct 06;5(10):e136 [FREE Full text] [doi: 10.2196/mhealth.6640] [Medline: 28986340]
7. Labrique AB, Vasudevan L, Kochi E, Fabricant R, Mehl G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. Glob Health Sci Pract 2013 Aug;1(2):160-171 [FREE Full text] [doi: 10.9745/GHSP-D-13-00031] [Medline: 25276529]
8. Eckman M, Gorski I, Mehta K. Leveraging design thinking to build sustainable mobile health systems. J Med Eng Technol 2016;40(7-8):422-430. [doi: 10.1080/03091902.2016.1218560] [Medline: 27535325]
9. Aranda-Jan CB, Mohutsiwa-Dibe N, Loukanova S. Systematic review on what works, what does not work and why of implementation of mobile health (mHealth) projects in Africa. BMC Public Health 2014 Feb 21;14:188 [FREE Full text] [doi: 10.1186/1471-2458-14-188] [Medline: 24555733]
10. Palepu K, Khanna T. Institutional Voids and Policy Challenges in Emerging Markets. The Brown Journal of World Affairs 1998;5(1):71-78.
11. Khanna T, Palepu K. Winning in Emerging Markets: A Road Map for Strategy and Execution. Boston, MA: Harvard Business Review Press; 2010.
12. Wang MD, Khanna R, Najafi N. Characterizing the Source of Text in Electronic Health Record Progress Notes. JAMA Intern Med 2017 Aug 01;177(8):1212-1213 [FREE Full text] [doi: 10.1001/jamainternmed.2017.1548] [Medline: 28558106]
13. Singletary B, Patel N, Heslin M. Patient Perceptions About Their Physician in 2 Words: The Good, the Bad, and the Ugly. JAMA Surg 2017 Dec 01;152(12):1169-1170. [doi: 10.1001/jamasurg.2017.3851] [Medline: 29049444]
14. ESSENCE. Sangath. URL: https://sangath.in/essence-3/ [accessed 2021-05-29]
15. Modi D, Dholakia N, Gopalan R, Venkatraman S, Dave K, Shah S, et al. mHealth intervention "ImTeCHO" to improve delivery of maternal, neonatal, and child care services-A cluster-randomized trial in tribal areas of Gujarat, India. PLoS Med 2019 Oct;16(10):e1002939 [FREE Full text] [doi: 10.1371/journal.pmed.1002939] [Medline: 31647821]
16. Scholten H, Granic I. Use of the Principles of Design Thinking to Address Limitations of Digital Mental Health Interventions for Youth: Viewpoint. J Med Internet Res 2019 Jan 14;21(1):e11528 [FREE Full text] [doi: 10.2196/11528] [Medline: 31344671]
17. Brown T. Design thinking. Harv Bus Rev 2008 Jun;86(6):84-92, 141. [Medline: 18605031]
18. Kazi DS, Greenough PG, Madhok R, Heerboth A, Shaikh A, Leaning J, et al. Using mobile technology to optimize disease surveillance and healthcare delivery at mass gatherings: a case study from India's Kumbh Mela. J Public Health (Oxf) 2017 Sep 01;39(3):616-624 [FREE Full text] [doi: 10.1093/pubmed/fdw091] [Medline: 27694349]
19. Agarwal S, Perry HB, Long L, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]
20. Malan JL, Newberry J. Human-Centred Design In The Field. UNICEF. URL: https://www.unicef.org/innovation/media/12111/file [accessed 2021-10-28]
21. Digital Implementation Investment Guide (DIIG): Integrating Digital Interventions into Health Programmes. World Health Organization. 2020. URL: https://www.who.int/publications/i/item/9789240010567 [accessed 2021-10-28]
22. Human-Centered Design 4 Health. UNICEF. URL: https://www.hcd4health.org/hcd-health-homepage [accessed 2020-11-01]
23. Kumbh Mela. Encyclopedia Britannica. URL: https://www.britannica.com/topic/Kumbh-Mela [accessed 2020-10-29]
24. Balsari S, Vemulapalli P, Gofine M, Oswal K, Merchant R, Saunik S, et al. A retrospective analysis of hypertension screening at a mass gathering in India: implications for non-communicable disease control strategies. J Hum Hypertens 2017 Nov;31(11):750-753. [doi: 10.1038/jhh.2017.54] [Medline: 28726795]
25. 2011 Census Data. Census India. URL: http://censusindia.gov.in/2011-Common/CensusData2011.html [accessed 2020-11-01]

26. Kumbh Mela: A million man dip. The Christian Science Monitor. 2013. URL: https://www.csmonitor.com/World/Global-News/2013/0116/Kumbh-Mela-A-million-man-dip [accessed 2020-10-29]
27. Stage set for Simhastha Kumbh Mela. Deccan Herald. URL: https://www.deccanherald.com/content/489106/stage-set-simhastha-kumbh-mela.html [accessed 2021-11-12]
28. Harris G. Deadly Stampede at Hindu Festival That Draws Millions. The New York Times. 2013. URL: https://www.nytimes.com/2013/02/12/world/asia/deadly-stampede-at-hindu-festival-that-draws-millions.html [accessed 2020-10-29]
29. Memish ZA, Zumla A, Alhakeem RF, Assiri A, Turkestani A, Al Harby KD, et al. Hajj: infectious disease surveillance and control. The Lancet 2014 Jun 14;383(9934):2073-2082 [FREE Full text] [doi: 10.1016/S0140-6736(14)60381-0] [Medline: 24857703]
30. Vortmann M, Balsari S, Holman SR, Greenough PG. Water, Sanitation, and Hygiene at the World's Largest Mass Gathering. Curr Infect Dis Rep 2015 Feb;17(2):461. [doi: 10.1007/s11908-015-0461-1] [Medline: 25783442]
31. Kumar GS, Kar SS, Jain A. Health and environmental sanitation in India: Issues for prioritizing control strategies. Indian J Occup Environ Med 2011 Sep;15(3):93-96 [FREE Full text] [doi: 10.4103/0019-5278.93196] [Medline: 22412284]
32. Banks A. Religious Fairs and Festivals in India. The Lancet 1961 Jan 21;277(7169):162-163. [doi: 10.1016/s0140-6736(61)91330-7] [Medline: 13686602]
33. Balsari S, Greenough PG, Kazi D, Heerboth A, Dwivedi S, Leaning J. Public health aspects of the world's largest mass gathering: The 2013 Kumbh Mela in Allahabad, India. J Public Health Policy 2016 Dec;37(4):411-427. [doi: 10.1057/s41271-016-0034-z] [Medline: 28202925]
34. IDEO. The Field Guide to Human-Centered Design. San Francisco, CA: IDEO; 2015.
35. Roberts JP, Fisher TR, Trowbridge MJ, Bent C. A design thinking framework for healthcare management and innovation. Healthc (Amst) 2016 Mar;4(1):11-14. [doi: 10.1016/j.hjdsi.2015.12.002] [Medline: 27001093]
36. Odendaal W, Anstey Watkins J, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3:CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
37. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12992-017-0275-z] [Medline: 28756767]
38. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
39. Art. 5 GDPR: Principles relating to processing of personal data. General Data Protection Regulation. URL: https://gdpr-info.eu/art-5-gdpr/ [accessed 2020-10-29]
40. Wierda E, Blok S, Somsen GA, van der Velde ET, Tulevski II, Stavrov B, et al. Protecting patient privacy in digital health technology: the Dutch m-Health infrastructure of Hartwacht as a learning case. BMJ Innov 2020 Jul 02;6(4):170-176. [doi: 10.1136/bmjinnov-2019-000399]
41. Jana A, Basu R. Glob Health Res Policy 2017;2:6 [FREE Full text] [doi: 10.1186/s41256-017-0026-y] [Medline: 29202074]
42. Mehndiratta A, Sharma S, Gupta NP, Sankar MJ, Cluzeau F. Adapting clinical guidelines in India-a pragmatic approach. BMJ 2017 Nov 17;359:j5147 [FREE Full text] [doi: 10.1136/bmj.j5147] [Medline: 29150419]
43. Das J, Holla A, Das V, Mohanan M, Tabak D, Chan B. In urban and rural India, a standardized patient study showed low levels of provider training and huge quality gaps. Health Aff (Millwood) 2012 Dec;31(12):2774-2784 [FREE Full text] [doi: 10.1377/hlthaff.2011.1356] [Medline: 23213162]
44. Bhatia A, Matthan R, Khanna T, Balsari S. Regulatory Sandboxes: A Cure for mHealth Pilotitis? J Med Internet Res 2020 Sep 15;22(9):e21276 [FREE Full text] [doi: 10.2196/21276] [Medline: 32763889]

Abbreviations

MRN: medical record number
NDHM: National Digital Health Mission
OPD: outpatient department

Edited by R Kukafka; submitted 15.02.21; peer-reviewed by F Lami, F Ghezelbash, N Muinga, L Ortiz-Comino; comments to author 12.05.21; revised version received 23.08.21; accepted 25.09.21; published 10.01.22.

Please cite as:
Shaikh A, Bhatia A, Yadav G, Hora S, Won C, Shankar M, Heerboth A, Vemulapalli P, Navalkar P, Oswal K, Heaton C, Saunik S, Khanna T, Balsari S
Applying Human-Centered Design Principles to Digital Syndromic Surveillance at a Mass Gathering in India: Viewpoint
J Med Internet Res 2022;24(1):e27952
URL: https://www.jmir.org/2022/1/e27952
doi:10.2196/27952
PMID:35006088

Viewpoint

Fine Detection of Human Motion During Activities of Daily Living as a Clinical Indicator for the Detection and Early Treatment of Chronic Diseases: The E-Mob Project

David Thivel[1*], PhD; Alice Corteval[2*], PharmD; Jean-Marie Favreau[1*], PhD; Emmanuel Bergeret[1*], PhD; Ludovic Samalin[3*], MD, PhD; Frédéric Costes[3*], MD, PhD; Farouk Toumani[1*], PhD; Christian Dualé[3*], MD, PhD; Bruno Pereira[3*], PhD; Alain Eschalier[3*], MD, PhD; Nicole Fearnbach[4*], PhD; Martine Duclos[3*], MD, PhD; Anne Tournadre[3*], MD, PhD

[1]Clermont Auvergne University, Aubiere, France

[2]Analgesia Foundation, Clermont-Ferrand, France

[3]Clermont Ferrand University Hospital, Clermont-Ferrand, France

[4]Pennington Biomedical Research Center, Baton Rouge, LA, United States

[*]all authors contributed equally

Corresponding Author:
David Thivel, PhD
Clermont Auvergne University
3 rue de la chabarde
Aubiere, 63170
France
Phone: 33 0770398975
Email: david.thivel@uca.fr

Abstract

Methods to measure physical activity and sedentary behaviors typically quantify the amount of time devoted to these activities. Among patients with chronic diseases, these methods can provide interesting behavioral information, but generally do not capture detailed body motion and fine movement behaviors. Fine detection of motion may provide additional information about functional decline that is of clinical interest in chronic diseases. This perspective paper highlights the need for more developed and sophisticated tools to better identify and track the decomposition, structuration, and sequencing of the daily movements of humans. The primary goal is to provide a reliable and useful clinical diagnostic and predictive indicator of the stage and evolution of chronic diseases, in order to prevent related comorbidities and complications among patients.

(J Med Internet Res 2022;24(1):e32362) doi:10.2196/32362

KEYWORDS

indicator; fine body motion; movement behaviors; decomposition; structuration; sequencing

Introduction

The alarming increases in physical inactivity and sedentary behaviors that accompanied societal development have favored the progression of chronic diseases. More patients now require ongoing medical attention and have limitations on activities of daily living, leading to reductions in life expectancy and health span [1]. Public health recommendations today advocate for the adoption of a minimal amount of moderate-to-vigorous physical activity (MVPA) each day, with a limited time devoted to sedentary activities (eg, daily sitting and screen time) [2]. While scientific societies and public health agencies long focused on the promotion of physical activity (PA), they also now include sedentary guidelines for overall health, with sedentary time independent of an individual's PA level [2]. According to Ekelund et al, 65 minutes per day of MVPA may be required to counteract the negative effects of 6 to 7 hours of daily sitting time on overall health [3]. These results clearly highlight the need for the joint promotion of both PA and reduced sedentary time. While these epidemiological results and associated recommendations are based on the general population, data regarding the relationship between the PA/sedentary profile of patients with chronic diseases and the bidirectional associations with disease progression and comorbidity are insufficient.

Methods to measure PA and sedentary behaviors typically quantify the amount of time devoted to these activities, estimating their intensity and frequency. Estimates of energetic costs (in metabolic equivalents of task [METs]) and energy expenditure are also available from most methods. The large majority of tools and technologies (eg, interviews, questionnaires, and wearable devices) that are available for tracking PA/sedentary behavior use these metrics. There are numerous reports of the validity and reliability/reproducibility of these tools among patients with various diseases, but the results are relatively poor to modest [4-6]. Well-conducted reviews (with systematic and meta-analytic approaches) have discussed in detail the inherent methodological limitations of the tools and devices commonly used to assess PA and sedentary behavior [7-9].

Moreover, quantifying PA and sedentary behavior among patients with chronic diseases can provide interesting behavioral information, but generally does not capture detailed and fine motions. Fine detection of motion may provide additional information about functional decline that is of clinical interest in chronic diseases. Of interest is the identification and tracking of the decomposition, structuration, and sequencing of humans' daily movements (meaning here the identification of the involved limbs, their respective contribution, and the temporal order of implication), above the simple quantification of PA and sedentary behavior. These metrics could provide a reliable and useful clinical diagnostic and predictive indicator (as an early sign) of the stage and evolution of chronic diseases and related comorbidities and complications. In the longer term, advancement in this direction could potentially influence treatment strategies (posology, timing/chronobiology, and nature of the treatment).

Body Movements' Decomposition Above PA and Sedentary Behavior Quantification

A better evaluation or capture of an individual's singular motion pattern or movement construction, from a musculoskeletal point of view, may allow practitioners to anticipate or track the evolution of some chronic diseases. In their recent work, Chevance et al proposed such an approach, showing that the anticipation and detection of the early signs of individuals' movement changes, as an indicator of subsequent critical functional gain or loss, need to be considered as "early warning signals for sudden behavioral changes" [10]. Beyond PA and overall human movements themselves, early findings suggest that sudden gain or loss in complex systems could be predicted through early warning signals [11,12], such as slight changes or fluctuations in human motion and movement patterns. On application to movement behaviors, the early anticipation of imminent body motion disruption or early detection of the first signs of fluctuations might represent a potentially reliable signal for delivering "just-in-time" interventions [13]. In that sense, Chevance et al observed that in adults with obesity, fluctuations in walking patterns were associated with the subsequent occurrence of behavioral losses in the following days, clearly demonstrating the need to develop new accessible methods to properly detect such early signals [10]. This need is clearly illustrated by high-quality clinical studies.

Clinical Evidence

Krieger et al, for instance, clearly noted the need to better identify subprocesses of movement execution in patients with schizophrenia, also highlighting that some neuroleptic treatments have negative side effects such as the slowing of motor execution [14]. The slowing of movement is nonperceptible and not captured with available activity trackers. Obesity has also been shown to affect patients' body motion and movement patterns, limiting their upper body range of motion during daily activities [15] or patterns of gait through mediolateral adaptations of their gravity center [16]. Once more, while this remains difficult to track in free-living conditions, the evolution of such movement patterns could be of great interest in the clinical care of these patients. In their work, Oubre et al also underlined that movement decomposition captures the core features of ataxia and may be useful for objective, precise, and frequent assessment of ataxia in home and clinic environments [17]. While not exhaustive, these examples clearly point out the urgent need to develop new strategies and tools to better track and catch the fine-grained evolution of patients' body movements in addition to the quantification of habitual PA and sedentary time.

Need for New Technologies

Importantly, such a fine-grained clinical exploration of daily motion does not negate the utility and interest for the activity trackers developed to date, but calls for a new and deeper integration and understanding of their signals and sensing capabilities. Commercialized trackers have shown satisfactory acceptability in capturing the daily routine of individuals. We should build on this platform, developing more complex and sophisticated algorithms to better identify and refine human movement patterns [18]. Such a process has been initiated through the development of human activity recognition (HAR) that uses wearable motion sensors, for which a high level of accuracy in predicting activities has been reported [19,20]. These sensors and algorithms have been shown to be valid and reliable among healthy individuals, but lack sensitivity to properly classify human movement in clinical patients, particularly patients presenting motor and gait impairments [9,21]. Moving forward, there is a need to optimize and validate these existing algorithms among patients with chronic diseases [5]. Previously published studies have shown that the validity of existing algorithms to discriminate sedentary behavior from standing and dynamic body behaviors and activities varies and mainly relies on explorations with reduced sample sizes [6,22-29]. While van Dijk-Huisman et al recently proposed an optimized PA classification based on signals from classical accelerometers that reliably classify sedentary and dynamic activities and detect postural transition among hospitalized patients [30], these approaches need to be developed for free-living conditions. Teams of clinicians and engineers should collaborate on technological innovations in measurement tools to fit patient characteristics and treatment plan needs.

Perspectives and New Technological Challenges

The vast majority of publications and commercial activity tracking solutions rely on the recognition of less than 10 activities (walking, climbing stairs, cycling, etc) [31]. This semantic description does not reflect the fine movements that would be informative in clinical settings. Moreover, they mainly focus on movements that are less typical in day-to-day life (eg, sports activity sessions), and do not capture the whole spectrum of nonexercise activities and finer motions that characterize our contemporary sedentary lifestyle. Understanding the quality (speed and trajectory) of fine movements, as well as the alteration or improvement in movement sequencing or frequency, would be more informative for describing the onset and progression of chronic diseases.

Several challenges are therefore identified and need to be addressed. First, it will be necessary for sensor devices to meet the following requirements: (1) avoid commercial devices or algorithms that do not retain data in the event of very little movement, as these are precisely the phenotypes we are interested in; (2) assume an autonomy of the order of 1 week, to make realistic the use of the device within the framework of a follow-up at home for patients, by ensuring the capture and the treatment of the signals at a frequency of the order of 50 Hz; (3) use sensors at multiple locations on the body in order to significantly capture fine movements related to sedentary behaviors and activities of daily living; and (4) have sensors with representative measurements (such as an accelerometer and gyroscope). The use of very lightweight neural networks, which work on a network personalization approach to customize the training to each patient, seems to be the most promising approach to date.

One of the primary approaches to reduce the energy cost of storing and transmitting data from sensors will be to move the data processing as close as possible to the sensors. In particular, we plan to use tools, such as Tensorflow light, to embed this processing in low-energy sensor devices (much lower than the power of a smartphone, for example). This will involve designing deep neural networks that take into account this distributed computation, where only partial data will come back from each device in order to finish the processing. We also plan to propose unconnected devices, in order to minimize the electronic complexity of the devices, with the transmission taking place over a wired connection when the devices are recharged. The charging and final data collection device could be integrated into a smart home-type environment, which would be a favorable and ecological solution, with little modification to the habits and facilities of the patients who will be equipped.

Such an approach would first need to define and validate a representative taxonomy of activities, from a musculoskeletal point of view, which could be integrated into a reliable and elaborate processing chain. The first approach would rely on a hierarchical taxonomy able to deeply detail the structure, substructure, and finesse of the detected semantics, and to identify and recognize any potential improvement or degradation linked with the evolution of the pathology.

Discussion: The E-Mob Project

In the above context and as part of the 2020-2025 scientific priorities that include “Human Mobility and Health” (I-site CAP 20-25, third challenge), the University Clermont Auvergne (Clermont-Ferrand, France) gathered an interdisciplinary group of experts (composed of physicians, physiologists, methodologists, biostatisticians, experts in energy metabolism, as well as computer scientists and developers) with the objective to elaborate and conduct a whole research program (the E-Mob project) aiming at (1) improving our technological abilities to precisely and accurately identify fine human body movements that might be relevant and informative when it comes to chronic diseases and (2) determining potential specific “digital movement signatures” that could help predict and follow the evolution of some chronic diseases and serve as a reliable connected support when it comes to treatment strategies.

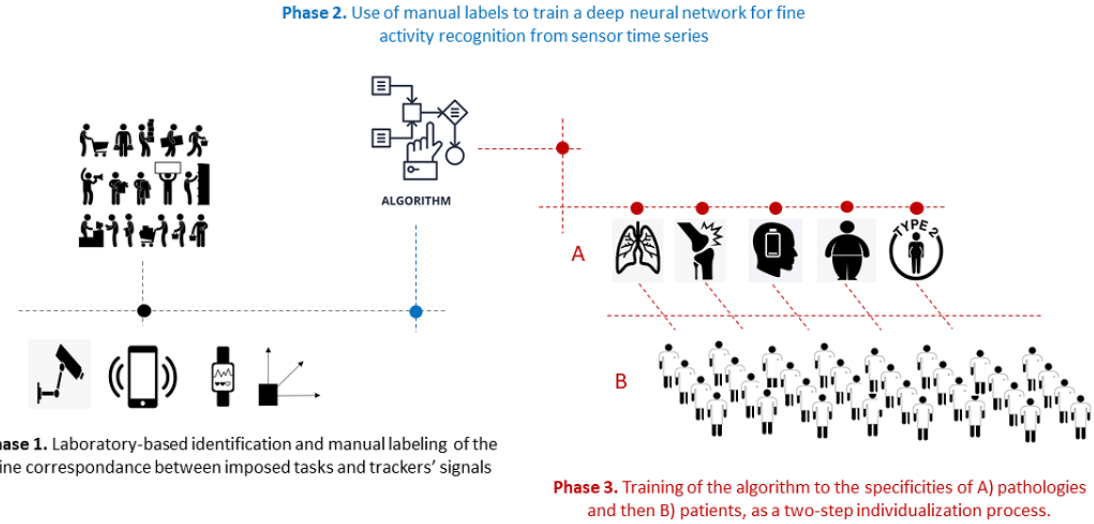
Briefly, the main idea of the E-Mob project is to propose the integration of the evaluation of detailed and fine human motions as a real clinical indicator. In that sense, we aimed at developing an in-clinic high-resolution human motion setting to perform regular deep and fine evaluation of patient mobility, and to develop an original sensor device and a dedicated algorithmic process to regularly assess body motion while engaged in PA and sedentary behavior in free-living conditions. The results obtained during this free-living evaluation would be directly uploaded from home to the data center to be analyzed and would help the clinical staff (physicians and physiotherapists, as well as nurses who are most of the time the first interlocutor of patients) determine whether the evolution of the activity pattern of the patient is an early warning signal of disease.

The E-Mob project will then develop new personalized HAR algorithms using artificial intelligence and machine learning, which will permit, from accelerometer signals, a qualitative analysis including the executive realization of movement and the identification of a pattern of activity. These algorithms should be personalized and should consider activities of very low intensity, the movement itself (function and performance), its sequence, breaks in sedentary behavior, and specificities to the underlying disease. Ultimately, this project aims to develop a continuous e-strategy to improve our personalized medicine approach. The need for such an approach based on daily human motion as a crucial early clinical indicator, has been emphasized during the COVID-19 pandemic, with clear links between the reduction of healthy movement behaviors and the progression of diseases, particularly metabolic and mental ones.

Phase 1 of the project will rely on the identification of specific signals from activity trackers, using a preliminary experiment asking healthy adult individuals to perform preselected and determined movements. This first phase will allow us to isolate the exact signals corresponding to the substages of these movements. From there, new algorithms will be developed in Phase 2. As a 2-step individualization process (Phase 3), these new algorithms will then be tested and trained to the specificities of several chronic diseases and to individual patients in order to identify a digital signature predictive of the state and course of the diseases. We will also capture the development of

multimorbidity, focused on the trajectory of each pathology and of each patient. These evaluations will then be replicated during a 4-year longitudinal study. These cohorts will include patients with inflammatory rheumatic diseases (rheumatoid arthritis and spondyloarthritis), and knee and hip osteoarthritis, as well as patients with chronic pain, obesity, type 2 diabetes, chronic obstructive pulmonary disease, major depressive disorders, bipolar disorders, and other mental health conditions. Figure 1 illustrates the steps of the E-Mob project.

Figure 1. Schematic flow chart of the E-Mob project from Phase 1 (detailed identification of signals from different trackers, specific to predetermined body movements) to Phase 2 (training of the deep neural network) and then Phase 3 (2-step individualization process).



Clinical Implications

There is a clear transition toward the development of more personalized medicine and individualized care, with an essential role for adapted e-technologies. The gross and fine tracking of body motions and the characteristics of these movement patterns over time will have direct clinical implications. The E-Mob solution will allow for the early detection of abnormal movement patterns and fine motions, and allow for targeted care strategies to minimize disease progression. This approach should favor the anticipation of clinical complications and prevent the loss of autonomy or the increased risk of falls associated with chronic diseases. In addition, it may help address the viscous circle of inactivity, isolation, or a hypotonic state in patients with mental and psychological disorders.

This fine tracking of our patients' daily movements might not only provide information regarding the degradation of their clinical condition but also help evaluate the effect of treatment. Overall, these approaches could aid in decisions regarding clinical care and intervention. Adapted PA programs and therapeutic education have been shown to be appropriate and necessary for most chronic diseases. More finely identifying each patient's physical limitations, weaknesses, and needs will help practitioners (physiotherapists, adapted educators, and nurses) directly tailor PA prescriptions. Self-monitoring has a positive impact on movement behaviors in both healthy individuals and those with chronic diseases, leading to increased PA and reduced sedentary time, which can translate to long-term beneficial effects.

Conclusion

Clinical care teams today rely more on e-technologies to develop and improve individualized and personalized medicine. The E-Mob project proposes to integrate such technologies in not only the short-term treatment of patients but also daily life and routine. This approach would allow practitioners to anticipate any evolution of chronic conditions, particularly using premature detection of abnormal fine body motions as an early clinical indicator to consider in care.

Acknowledgments

The authors thank the I-Site project of Clermont Auvergne University for supporting the E-Mob project.

Conflicts of Interest

None declared.

References

1. Min C, Yoo DM, Wee JH, Lee H, Byun SH, Choi HG. Mortality and cause of death in physical activity and insufficient physical activity participants: a longitudinal follow-up study using a national health screening cohort. BMC Public Health 2020 Sep 29;20(1):1469 [FREE Full text] [doi: 10.1186/s12889-020-09564-x] [Medline: 32993602]

2. Bull FC, Al-Ansari SS, Biddle S, Borodulin K, Buman MP, Cardon G, et al. World Health Organization 2020 guidelines on physical activity and sedentary behaviour. Br J Sports Med 2020 Dec 25;54(24):1451-1462 [FREE Full text] [doi: 10.1136/bjsports-2020-102955] [Medline: 33239350]
3. Ekelund U, Tarp J, Steene-Johannessen J, Hansen BH, Jefferis B, Fagerland MW, et al. Dose-response associations between accelerometry measured physical activity and sedentary time and all cause mortality: systematic review and harmonised meta-analysis. BMJ 2019 Aug 21;366:l4570 [FREE Full text] [doi: 10.1136/bmj.l4570] [Medline: 31434697]
4. Fazio S, Stocking J, Kuhn B, Doroy A, Blackmon E, Young HM, et al. How much do hospitalized adults move? A systematic review and meta-analysis. Appl Nurs Res 2020 Feb;51:151189 [FREE Full text] [doi: 10.1016/j.apnr.2019.151189] [Medline: 31672262]
5. Lim SER, Ibrahim K, Sayer AA, Roberts HC. Assessment of Physical Activity of Hospitalised Older Adults: A Systematic Review. J Nutr Health Aging 2018 May 19;22(3):377-386. [doi: 10.1007/s12603-017-0931-2] [Medline: 29484351]
6. Ainsworth B, Cahalin L, Buman M, Ross R. The current state of physical activity assessment tools. Prog Cardiovasc Dis 2015;57(4):387-395. [doi: 10.1016/j.pcad.2014.10.005] [Medline: 25446555]
7. Brickwood K, Watson G, O'Brien J, Williams AD. Consumer-Based Wearable Activity Trackers Increase Physical Activity Participation: Systematic Review and Meta-Analysis. JMIR Mhealth Uhealth 2019 Apr 12;7(4):e11819 [FREE Full text] [doi: 10.2196/11819] [Medline: 30977740]
8. Fuller D, Colwell E, Low J, Orychock K, Tobin MA, Simango B, et al. Reliability and Validity of Commercially Available Wearable Devices for Measuring Steps, Energy Expenditure, and Heart Rate: Systematic Review. JMIR Mhealth Uhealth 2020 Sep 08;8(9):e18694 [FREE Full text] [doi: 10.2196/18694] [Medline: 32897239]
9. Evenson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer-wearable activity trackers. Int J Behav Nutr Phys Act 2015 Dec 18;12(1):159 [FREE Full text] [doi: 10.1186/s12966-015-0314-1] [Medline: 26684758]
10. Chevance G, Baretta D, Heino M, Perski O, Olthof M, Klasnja P, et al. Characterizing and predicting person-specific, day-to-day, fluctuations in walking behavior. PLoS One 2021 May 14;16(5):e0251659 [FREE Full text] [doi: 10.1371/journal.pone.0251659] [Medline: 33989338]
11. Scheffer M, Bascompte J, Brock WA, Brovkin V, Carpenter SR, Dakos V, et al. Early-warning signals for critical transitions. Nature 2009 Sep 03;461(7260):53-59. [doi: 10.1038/nature08227] [Medline: 19727193]
12. Haken H. An Introduction Nonequilibrium Phase Transitions and Self-Organization in Physics, Chemistry and Biology. Berlin, Heidelberg: Springer-Verlag; 1978.
13. Hardeman W, Houghton J, Lane K, Jones A, Naughton F. A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. Int J Behav Nutr Phys Act 2019 Apr 03;16(1):31 [FREE Full text] [doi: 10.1186/s12966-019-0792-7] [Medline: 30943983]
14. Krieger S, Lis S, Gallhofer B. Cognitive subprocesses and schizophrenia. A. Reaction-time decomposition. Acta Psychiatr Scand Suppl 2001(408):18-27. [doi: 10.1034/j.1600-0447.2001.104s408018.x] [Medline: 11730070]
15. Cau N, Cimolin V, Brugliera L, Ventura G, Galli M, Capodaglio P. Range of motion limitations of the upper body in obese female workers. Med Lav 2017 Dec 14;108(6):455-465. [doi: 10.23749/mdl.v108i6.6339] [Medline: 29240042]
16. Peyrot N, Thivel D, Isacco L, Morin J, Duche P, Belli A. Do mechanical gait parameters explain the higher metabolic cost of walking in obese adolescents? J Appl Physiol (1985) 2009 Jun;106(6):1763-1770 [FREE Full text] [doi: 10.1152/japplphysiol.91240.2008] [Medline: 19246657]
17. Oubre B, Daneault J, Whritenour K, Khan NC, Stephen CD, Schmahmann JD, et al. Decomposition of Reaching Movements Enables Detection and Measurement of Ataxia. Cerebellum 2021 Dec 02;20(6):811-822 [FREE Full text] [doi: 10.1007/s12311-021-01247-6] [Medline: 33651372]
18. Debache I, Jeantet L, Chevallier D, Bergouignan A, Sueur C. A Lean and Performant Hierarchical Model for Human Activity Recognition Using Body-Mounted Sensors. Sensors (Basel) 2020 May 29;20(11):3090 [FREE Full text] [doi: 10.3390/s20113090] [Medline: 32486068]
19. Awais M, Mellone S, Chiari L. Physical activity classification meets daily life: Review on existing methodologies and open challenges. Annu Int Conf IEEE Eng Med Biol Soc 2015;2015:5050-5053. [doi: 10.1109/EMBC.2015.7319526] [Medline: 26737426]
20. Figo D, Diniz PC, Ferreira DR, Cardoso JMP. Preprocessing techniques for context recognition from accelerometer data. Pers Ubiquit Comput 2010 Mar 30;14(7):645-662. [doi: 10.1007/s00779-010-0293-9]
21. Chen K, Janz K, Zhu W, Brychta R. Redefining the Roles of Sensors in Objective Physical Activity Monitoring. Medicine & Science in Sports & Exercise 2012;44(1S):S13-S23. [doi: 10.1249/mss.0b013e3182399bc8]
22. Pedersen MM, Bodilsen AC, Petersen J, Beyer N, Andersen O, Lawson-Smith L, et al. Twenty-four-hour mobility during acute hospitalization in older medical patients. J Gerontol A Biol Sci Med Sci 2013 Mar 12;68(3):331-337. [doi: 10.1093/gerona/gls165] [Medline: 22972940]
23. Valkenet K, Veenhof C. Validity of three accelerometers to investigate lying, sitting, standing and walking. PLoS One 2019 May 23;14(5):e0217545 [FREE Full text] [doi: 10.1371/journal.pone.0217545] [Medline: 31120994]
24. Kramer SF, Cumming T, Churilov L, Bernhardt J. Measuring activity levels at an acute stroke ward: comparing observations to a device. Biomed Res Int 2013;2013:460482-460488 [FREE Full text] [doi: 10.1155/2013/460482] [Medline: 24282815]

25. Baldwin CE, Johnston KN, Rowlands AV, Williams MT. Physical Activity of ICU Survivors during Acute Admission: Agreement of the activPAL with Observation. Physiother Can 2018 Feb;70(1):57-63 [FREE Full text] [doi: 10.3138/ptc.2016-61] [Medline: 29434419]
26. Brown CJ, Roth DL, Allman RM. Validation of use of wireless monitors to measure levels of mobility during hospitalization. J Rehabil Res Dev 2008 Dec 01;45(4):551-558 [FREE Full text] [doi: 10.1682/jrrd.2007.06.0086] [Medline: 18712640]
27. Anderson JL, Green AJ, Yoward LS, Hall HK. Validity and reliability of accelerometry in identification of lying, sitting, standing or purposeful activity in adult hospital inpatients recovering from acute or critical illness: a systematic review. Clin Rehabil 2018 Feb 14;32(2):233-242. [doi: 10.1177/0269215517724850] [Medline: 28805075]
28. Taraldsen K, Askim T, Sletvold O, Einarsen E, Bjåstad KG, Indredavik B, et al. Evaluation of a body-worn sensor system to measure physical activity in older people with impaired function. Phys Ther 2011 Feb;91(2):277-285. [doi: 10.2522/ptj.20100159] [Medline: 21212377]
29. Anderson JL, Yoward LS, Green AJ. A study investigating the validity of an accelerometer in quantification of step count in adult hospital inpatients recovering from critical illness. Clin Rehabil 2019 May 14;33(5):936-942. [doi: 10.1177/0269215519829893] [Medline: 30764647]
30. van Dijk-Huisman HC, Bijnens W, Senden R, Essers JMN, Meijer K, Aarts J, et al. Optimization and Validation of a Classification Algorithm for Assessment of Physical Activity in Hospitalized Patients. Sensors (Basel) 2021 Feb 27;21(5):1652 [FREE Full text] [doi: 10.3390/s21051652] [Medline: 33673447]
31. Sansano E, Montoliu R, Belmonte Fernández Ó. A study of deep neural networks for human activity recognition. Computational Intelligence 2020 Mar 27;36(3):1113-1139. [doi: 10.1111/coin.12318]

Abbreviations

PA: physical activity
MVPA: moderate-to-vigorous physical activity
HAR: human activity recognition

Edited by R Kukafka; submitted 24.07.21; peer-reviewed by G Dermody, MDG Pimentel, A Louren, P Li; comments to author 23.09.21; revised version received 24.10.21; accepted 29.10.21; published 14.01.22.

Please cite as:
Thivel D, Corteval A, Favreau JM, Bergeret E, Samalin L, Costes F, Toumani F, Dualé C, Pereira B, Eschalier A, Fearnbach N, Duclos M, Tournadre A
Fine Detection of Human Motion During Activities of Daily Living as a Clinical Indicator for the Detection and Early Treatment of Chronic Diseases: The E-Mob Project
J Med Internet Res 2022;24(1):e32362
URL: https://www.jmir.org/2022/1/e32362
doi:10.2196/32362
PMID:35029537

Viewpoint

The Role of Unobtrusive Home-Based Continuous Sensing in the Management of Postacute Sequelae of SARS CoV-2

Benjamin Harris Peterson Corman[1,2*], BA; Sritha Rajupet[3,4*], MD, MPH; Fan Ye[5*], PhD; Elinor Randi Schoenfeld[3*], PhD

[1]Renaissance School of Medicine, Stony Brook University, Stony Brook, NY, United States

[2]Program in Public Health, Stony Brook University, Stony Brook, NY, United States

[3]Department of Family, Population & Preventive Medicine, Renaissance School of Medicine, Stony Brook University, Stony Brook, NY, United States

[4]Department of Biomedical Informatics, Renaissance School of Medicine, Stony Brook University, Stony Brook, NY, United States

[5]Department of Electrical and Computer Engineering, College of Engineering and Applied Science, Stony Brook University, Stony Brook, NY, United States

[*]all authors contributed equally

Corresponding Author:
Fan Ye, PhD
Department of Electrical and Computer Engineering
College of Engineering and Applied Science
Stony Brook University
Light Engineering Building, Room 217
Stony Brook, NY, 11794-2350
United States
Phone: 1 631 632 8393
Fax: 1 631 632 8494
Email: fan.ye@stonybrook.edu

Abstract

Amid the COVID-19 pandemic, it has been reported that greater than 35% of patients with confirmed or suspected COVID-19 develop postacute sequelae of SARS CoV-2 (PASC). PASC is still a disease for which preliminary medical data are being collected—mostly measurements collected during hospital or clinical visits—and pathophysiological understanding is yet in its infancy. The disease is notable for its prevalence and its variable symptom presentation, and as such, management plans could be more holistically made if health care providers had access to unobtrusive home-based wearable and contactless continuous physiologic and physical sensor data. Such between-hospital or between-clinic data can quantitatively elucidate a majority of the temporal evolution of PASC symptoms. Although not universally of comparable accuracy to gold standard medical devices, home-deployed sensors offer great insights into the development and progression of PASC. Suitable sensors include those providing vital signs and activity measurements that correlate directly or by proxy to documented PASC symptoms. Such continuous, home-based data can give care providers contextualized information from which symptom exacerbation or relieving factors may be classified. Such data can also improve the collective academic understanding of PASC by providing temporally and activity-associated symptom cataloging. In this viewpoint, we make a case for the utilization of home-based continuous sensing that can serve as a foundation from which medical professionals and engineers may develop and pursue long-term mitigation strategies for PASC.

(J Med Internet Res 2022;24(1):e32713) doi:10.2196/32713

KEYWORDS

SARS CoV-2; COVID-19; post-acute sequelae of SARS CoV-2 (PASC); post-COVID; long COVID; continuous sensing; passive monitoring; wearable sensors; contactless sensors; vital sign monitoring

Introduction

Postacute Sequelae of SARS CoV-2

On March 11, 2020, the World Health Organization (WHO) declared COVID-19 a pandemic. Since then, the responsible virus, SARS-CoV-2, has spread across the international community, leading to the most pervasive and defining public health event of this generation. In the United States, as of this writing, there have been over 65 million cases and nearly 850,000 deaths [1]. Less than 1 year after COVID-19 was declared a pandemic, the Food and Drug Administration (FDA) issued Emergency Use Authorization for the Pfizer-BioNTech, Moderna, and Janssen COVID-19 vaccines [2]. Amid the ongoing US nationwide vaccination campaign, attention has turned to the allocation of resources to address the long-term health impacts of the pandemic.

The medical community has paid particular attention to those individuals whose COVID-19 symptoms never fully remitted, have since recurred, or have developed new symptoms thought to be related to having COVID-19. Current estimates are that upwards of 35% of COVID-19 patients fall into this group of individuals who have a disease now recognized formally as postacute sequelae of SARS CoV-2 (PASC) but is also colloquially being referred to as post-COVID, long COVID, long-haul COVID, postacute COVID-19, long-term effects of COVID, or chronic COVID [3,4]. This can translate to over 15 million PASC patients utilizing the health care system in the United States alone, with the numbers growing daily.

PASC occurs in any patient with a prior COVID-19 infection, yet the severity of the disease proves to be a poor predictor of prolonged or recurrent symptoms [3]. PASC patients present with a variety of multisystem symptoms that have made evaluation, diagnosis, and treatment more challenging. We present here a proposal for the widespread adoption of continuous sensing devices to provide contextualized surveillance data for patients' core vital signs and activities between medical encounters (eg, at home). Such continuous sensing and the subsequent raw data processing would provide quantitative means and insights into the symptom progression of PASC at home and between hospital or clinical visits. They will enable clinicians to holistically assess standards of care while simultaneously offering higher levels of operationalism to the management of PASC. We also provide suggestions on how to interpret such data in the context of PASC management. Evaluations of these data at the population level can serve as the foundation for the creation of predictive analytics tools that can be utilized in the management of all PASC patients. It must be stressed that, while we see substantial benefit in the use of continuous monitoring for PASC and chronic diseases in general, we do not intend to diminish the importance of current efforts to mitigate the acute effects of the pandemic, and such home-based sensing serves to complement and not replace the rigorous medical device–based data collection performed during hospital or clinical visits. Instead, our goal in this paper is to propose a framework from which medical professionals and engineers may pursue long-term mitigation strategies for PASC via home-based continuous sensing data in conjunction with ongoing medical and technical innovations and efforts in the acute care setting for COVID-19.

Background

A particular challenge with managing PASC is the fact that there has yet to be an established standard of care. Until recently, research on the topic could be described as sparse and grassroots, with patient reports of persistent symptoms, particularly those from disparate PASC support groups, driving clinical understanding [5]. PASC is now widely recognized by the medical community, with the National Institutes of Health (NIH) launching a broad-based initiative to understand the disease and the long-term health of this cohort of patients [6]. In a recent retrospective cohort study, it was found that 36.55% of all COVID-19 patients have at least one PASC-associated symptom between 3 and 6 months after their initial COVID-19 diagnosis [4]. Such an assessment captures the breadth of this public health issue and represents a substantial sustained burden on patients, families, communities, medical care systems, and the nation.

Despite the formal recognition of PASC, a notable lack of understanding of the pathophysiology of the syndrome remains. Although data from spirometry, electrocardiography, auscultation, stress tests, and other clinical measures can assist with developing a clinical picture of COVID-19's impact on the patient's lungs and heart, these tests represent point observations. Symptoms of PASC, however, have been found to wax and wane over time, necessitating continuous monitoring for accurate quantitative measures of symptoms [7]. Furthermore, it has been observed that some symptoms are exacerbated during the performance of specific daily tasks [3,8]. As these activities most often occur outside of the clinical setting, it is vitally important that data should be collected in a home (or home-like) environment to identify these potentially transient events.

Currently, in research and clinical studies, a popular method by which these data can be collected is by means of a symptom diary in which patients self-report changes in their physiological state. Such tools rely on retrospective reporting, making them prone to recall bias in addition to self-serving biases [9]. Diaries also pose a significant respondent burden, and as such, adherence to daily diary keeping is a significant consideration. In one study that required chronic pain participants to record 3 pain entries per day for 21 consecutive days, it was found that only 10.9% of traditional paper diary entries were recorded within the temporally defined compliance window set by the study coordinators, despite reported compliance on 90.5% of entries [10]. It has been found that there is limited evidence to suggest significant clinical utility in traditional symptom diaries, with one study concluding that among asthmatic children there was no statistically significant beneficial effect of keeping a symptom diary on day-to-day asthma control nor was there a decrease in the future risk of asthma control [11]. Although electronic diaries with measures designed to enhance use compliance have been shown to be associated with greater adherence to the temporal demands of keeping a symptom diary in some (but not all) contexts, there remains a significant user burden. As such, there remains nonuniversal adherence, and the utility of the collected data in informing clinical decisions with

beneficial outcomes is an area of ongoing research [9,10,12,13]. As such, there is a pressing need for a method of tracking symptoms that is standardized between patients, quantitative in its collection, and poses minimal patient or user burden. Sensing technologies, particularly those that are unobtrusive in nature, offer a promising solution to all 3 of these requirements.

Although there is a precedent for providing patients wearable sensors for vital sensing between hospital encounters, most notably in the form of Holter monitors for arrhythmia detection, their adoption has been far from widespread in the context of PASC, and the conspicuousness and complexity of these sensors limit their usability in all scenarios for extended periods [14,15]. A more user-friendly solution to long-term monitoring involves unobtrusive technologies that lend themselves to continuous passive use across a range of activities and whose output data is sufficiently robust to provide a reasonably accurate physiological picture.

Continuously monitoring sensors are already being implemented in the context of a multitude of chronic diseases including chronic obstructive pulmonary disease, asthma, congestive heart failure, and many more [16-19]. The use of such sensors in the context of PASC is an important tool for disease management. However, it has the additional benefit of providing substantial data for researchers to develop predictive analytics tools for PASC, making it an incredibly powerful tool for elucidating the relatively data-poor academic space of COVID-19 sequelae.

In the consumer market, there currently exist 2 primary means by which continuous vital sign data can be obtained: by way of wearable sensors and by way of contactless sensors. Wearable sensors require physical proximity (usually direct touch to the human body) between the user and the sensor [20]. This is the more pervasive sensing technology currently in the consumer market, with smartwatches, chest-strap heart rate monitors, glucometers, and smart clothing comprising some of the products in this family of technologies. Wearable sensors tend to be more portable and generally more accurate than contactless sensors but require users to charge and wear them regularly for accurate and consistent data collection.

Contactless sensors, by contrast, do not require physical proximity to a user [20]. Instead, they can collect data from a distance, without touching the human body. These sensors can function completely passively with no additional user efforts (ie, charging, wearing) beyond the initial setup. These sensors can be mounted to a wall or placed on a surface. However, while they have the benefit of operating with minimal user efforts, they lack portability. We believe it is with widespread utilization of a combination of wearable and contactless sensors that PASC pathophysiology can be best understood and management strategies can be best implemented. Although both sensor families offer complementary insights in disease tracking and managing and can further inform a clinician's understanding of the disease, it must be stressed that these sensors do not represent a replacement for existing gold standard instruments. Instead, sensor findings may prompt a clinician to perform further workup.

With regards to PASC diagnosis, there currently exists no single diagnostic criterion for PASC. The National Institute for Health and Care Excellence (NICE) in the United Kingdom defines "post-COVID-19 syndrome" as signs and symptoms that develop during or after an infection consistent with COVID-19, continue for more than 12 weeks, and are not explained by an alternative diagnosis [21]. In the United States, such a universal definition that can serve as the basis for diagnoses, even one as open-ended as that put forth by NICE, does not exist. Although not intended for the basis of a diagnostic criterion, the WHO recently put forth their own definition of the disease that was comparably vague in its description, describing a "condition [that] occurs in individuals with a history of probable or confirmed SARS-CoV-2 infection, usually 3 months from the onset of COVID-19 with symptoms that last for at least 2 months and cannot be explained by an alternative diagnosis" [22]. Similarly, the US Centers for Disease Control and Prevention (CDC) recognizes PASC as "a wide range of new, returning, or ongoing health problems people can experience four or more weeks after first being infected with the virus that causes COVID-19" [23]. As such, it is up to the clinician's discretion as to how and when to give a patient a diagnosis of PASC. Although continuous sensing will not change the lack of unified diagnostic criteria, it will provide a framework to quantitatively track signs and symptoms associated with PASC from the initial COVID-19 diagnosis (or suspected diagnosis) to the treatment of the acute phase and forward.

When treatment is indicated in the management of PASC, it is in the context of supportive management and symptom monitoring [24]. It is, therefore, important that medical providers are given data to inform the best medical treatment and to track patient progress. A symptom-specific means of continually monitoring a patient can provide quantitative measures to suggest symptom abatement or the efficacy of treatment.

Quantitatively Measurable PASC Symptoms

Symptom Presentation

The presentation of PASC is variable. Symptoms may be constant, or they may wax and wane. Generally, care providers obtain physiological measures of their patients at the time of clinical visits. Given that disease presentation is inconstant between patients and across time, having physiological measurements obtained with sensors between clinical visits would provide health care professionals a more holistic clinical picture.

Large-scale studies specifically designed to identify and catalog symptoms have been and are being conducted based on patient-reported symptoms [4,6,25-31]. Although each has some variability in the measured relative prevalence of symptoms, fatigue and shortness of breath are generally thought to be the most common symptoms [32]. Although further prevalence data are discrepant between studies, most common complaints of PASC involve the following systems or combinations thereof: respiratory, neurologic, psychiatric, metabolic, cardiovascular, gastrointestinal, musculoskeletal, and general well-being, as well as some less commonly identified other sequelae [3,25]. We provide in the following paragraphs a summary of

pathophysiologic burdens associated with PASC (either via direct causal association stemming from COVID-19 or due to preexisting conditions acting as a risk factor for the development of disease). We also offer potential sensor modalities whose implementation between medical encounters may directly or by proxy provide quantitative data to evaluate physiological changes associated with each respective symptom. A summary of PASC-associated symptoms and the potential sensing technologies that may quantify them is provided in Table 1 [29].

Table 1. Suggested sensor modalities to be used for the quantification of a selection of common postacute sequelae of SARS CoV-2 (PASC) symptoms.

Condition	Sensor modality
Acute coronary disease	Pose/motion monitor, activity monitor, pulse oximeter, heart rate
Anxiety	Sleep perturbations, activity monitor
Arrhythmia	Heart rate/rhythm monitor
Brain fog	Activity monitor
Changes in menstrual cycle	Heart rate monitor, temperature monitor
Chest pain	Activity monitor, pose/motion monitor
Constipation	Pose/motion monitor, activity monitor
Coughing	Microphone, pose/motion monitor
Depression	Sleep perturbations, activity monitor
Diabetes	Blood glucose monitor, insulin monitor, blood pressure monitor
Diarrhea	Pose/motion monitor, activity monitor
Fatigue	Sleep perturbations, activity monitor
Fever	Temperature monitor
Forgetfulness	Activity monitor
Gastroesophageal reflux disease	Heart rate, respiratory rate, pose/motion monitor
Headaches	Activity monitor
Heart failure	Heart rate/rhythm monitor
Hypoxemia	Pulse oximeter
Joint pain	Pose/motion monitor, heart rate, respiratory rate, activity monitor
Lightheadedness	Activity monitor, pose/motion monitor
Lipid metabolism disorders	Activity monitor, blood pressure monitor
Muscle weakness	Pose/motion monitor, heart rate, respiratory rate, activity monitor
Obesity	Weight tracking, activity monitor, blood pressure monitor
Olfactory perturbation	Activity monitor
Positional orthostatic tachycardia syndrome	Heart rate/rhythm monitor, pose/motion monitor
Posttraumatic stress disorder–associated symptoms	Sleep perturbations, activity monitor
Rash	Activity monitor, lesion identification
Shortness of breath	Pulse oximeter, respiratory rate
Stroke	Activity monitor, pose/motion monitor

Respiratory Manifestations

Given the propensity of COVID-19 to cause debilitating respiratory symptoms, the utility of continuous monitoring is particularly apparent in the context of measuring lung function. Shortness of breath is one of the most common presenting symptoms in both acute COVID-19 and PASC [25]. Existing personal sensor technologies are well-equipped to monitor basic respiratory signs including respiratory rate. A recent case study demonstrated a management plan involving continuously measuring respiration and pulse rates and daily step counts using a clothing-adhered sensor that utilized photoplethysmography, tri-axis accelerometers, and a dedicated sensor of respiratory effort. The sensor was used to monitor the at-home recovery of a COVID patient with several weeks of persistent respiratory symptoms [33]. A recent article presented the reliability of a novel skin-adhered strain sensor in measuring respiration rate and volume measurements [34]. The study demonstrated an ability to extrapolate volume and rate data from these sensors to coarsely predict spirometry results. Pulmonary function tests like spirometry are important in the context of PASC as both

restrictive and obstructive patterns of lung function have been reported among PASC patients [35].

Beyond shortness of breath, hypoxemia is a common respiratory complaint at 6 months following a positive COVID-19 diagnosis [25]. Pulse oximetry monitoring has long been a focus of personal wearable sensors. Fitbit, Apple, Garmin, and other smartwatch manufacturers have recently begun incorporating pulse oximetry as standard product features. These products are a complement to existing finger-based pulse oximeters whose utility as continuous monitoring devices are limited by size and inconvenience to the conduct of daily activities. Exertional dyspnea, the shortness of breath that occurs in the context of physical exertion, can be quantified with sensors that can track pulse oximetry in conjunction with markers of physical activity including heart rate, respiratory rate, or activity monitoring. Such a quantifiable measure of exertional dyspnea would allow for the early identification of patients in need of pulmonary or cardiac evaluation and the early initiation of therapeutic interventions [36-38].

Cough is another common PASC complaint. Current technologies exist that utilize continuous microphone-collected ambient sound to identify coughs using classification algorithms [39]. This offers a reasonable cost and resource-effective means of tracking coughing events across a spectrum of chronic diseases including PASC.

Cardiovascular Manifestations

In the context of PASC, cardiovascular symptoms and conditions of particular prevalence include chest pain, acute coronary disease, arrhythmias, and heart failure. Chest pain may be of cardiac, pulmonological, gastrointestinal, nervous, or musculoskeletal origin, but obtaining signs via proper continuous sensing can provide a means by which these etiologies can be identified and treatment can be most effectively directed and monitored.

Many of these heart conditions lend themselves particularly well to continuous sensing in ways that noncardiovascular chest pain may not [40-42]. Arrhythmias including sinus tachycardia, bradycardia, atrioventricular blocks, fibrillations, and flutters can all be picked up with a sufficiently sensitive heart rate monitor. These come standard in many commercial smartwatches, many of which now have primitive electrocardiogram capabilities that can theoretically detect symptoms of acute coronary disease before it manifests in myocardial infarction.

Additionally, some PASC patients report a condition called postural orthostatic tachycardia syndrome (POTS) [43]. Patients with POTS experience tachycardic symptoms in the context of postural change. As a result, a continuous, contextualized means of collecting postural and heart rate data would be integral to developing a full picture of the symptoms and their manifestations.

Psychiatric Manifestations

Many with PASC report psychiatric symptoms including feelings of anxiety and depression, with some patients having comorbid diagnoses of posttraumatic stress disorder. As with many of these symptom manifestations, there is not a single way by which continuous monitoring may be able to quantify these. There are, however, a multitude of modalities through which an adept care provider may make clinical recommendations. One such means by which diagnoses may be informed and treatment can be monitored is in the context of sleep disturbances. Many commercially available freestanding and smartphone-based contactless sensors as well as wearable devices including popular smartwatches track sleep quality, with some devices even offering the ability to stage sleep [44]. Assessing changes in quality, timing, duration, or staging of sleep can be useful in identifying these psychiatric symptoms and providing data to guide clinical intervention. Additionally, one particularly promising means by which psychiatric manifestations may be tracked is via activity and behavior changes. Sensors that can detect fine granularity changes in body pose, motion at seconds scale, and daily activities at minutes scale over time may elucidate psychiatric symptoms (though other symptom profiles may manifest in activity and behavior changes as well and thus need to be differentiated with high fidelity) [18,45]. These include wearable kinematic sensors or contactless depth sensors that can measure the coarse-grained 3-dimensional contours of surrounding objects, including the positional movement of users [46,47].

Neurologic Manifestations

Some of the symptoms commonly associated with PASC that are most prominently featured in the lay media are neurologic manifestations, particularly so-called “COVID-19 brain fog.” Many treatments have been purported to help “clear” brain fog, but its pathophysiology is, as of yet, poorly understood, and the home remedies suggested are far from universally effective [48]. Such a dearth of understanding necessitates the use of tools to actively monitor neurologic symptoms. Behavior monitoring, as proposed in the previous section, using kinematic sensors or depth-sensing devices may be effective at quantifying changes in neurologic status.

Neurologic manifestations of PASC that may manifest in behavior change in addition to brain fog include stroke, lightheadedness, memory problems, headache, and olfactory perturbations [25]. It has been reported that changes in blood pressure, body temperature, blood glucose, and blood oxygen saturation may also be associated with strokes. Thus, although not specific for neurologic involvement, tracking these vital signs may be useful in identifying the onset or exacerbation of neurologic dysfunction [49]. Lightheadedness is a symptom whose presenting signs may be particularly useful in the setting of contextualized data. For example, lightheadedness spells may be quantified based on altered movement (ie, swaying, or slow, purposive movement) in the context of rapid changes in body position as might occur with a sudden rise from a seated position. For other neurologic manifestations, namely forgetfulness, headaches, and olfactory perturbations, physiological changes may not be easily measured but might be inferred indirectly through sensing of certain motions (ie, hands massaging the face or skull in the setting of headaches). Though such inference of symptoms based on body motion and activity alone might be unreliable, the ability of such sensor data to be retroactively interpreted by both a physician and the

patient can help elucidate such basic metrics as date of symptom onset and frequency of recurrence as well as help identify exacerbating or relieving factors.

Metabolic Manifestations

Metabolism is an area of vital sensing that has historically been difficult to track. Although metabolic manifestations have been less commonly implicated than other systems in the context of PASC, patients have an excess burden of disorders of lipid metabolism, obesity, and diabetes mellitus. It is not yet clear to what extent these metabolic conditions in particular are associated with the disease pathology itself and to what extent underlying metabolic conditions predispose patients to PASC [50-52].

There currently exist products that, although not continuous sensing devices, claim to assess metabolic end products in exhaled air to give indications about the user's metabolic state [53]. Although an interesting proposal that demonstrates reasonable validity in initial studies, the product is yet to be widely tested and lacks FDA recognition.

Other potential sensing modalities include continuous glucose monitoring as has been widely used for those with diabetes mellitus. There have also been strides made towards developing flexible, adherent patches that utilize ultrasound to continuously measure blood pressure that can serve as a proxy for overall metabolic functioning [54,55]. Smartwatch-based blood pressure sensors are also in development, though the validity of such technologies has not been fully evaluated [56]. Slower-order changes like new or worsening obesity should be regularly monitored, but their slow progression may better fit the purview of interval monitoring, rather than continuous sensing. Smart scales and regular BMI measurements are better suited to quantify these trends.

Gastrointestinal Manifestations

Although there is a lack of published evidence about the effectiveness of apps and sensors in tracking and managing gastroenterological care, adopting technologies to enhance patient care in this field has been and remains an active area of academic focus [57]. Current sensors that may be able to identify gastrointestinal manifestations of PASC (including those of constipation and diarrhea) include activity or presence sensors that track a user's time in the bathroom. Gastroesophageal reflux disease, a reported manifestation of PASC, might be assessed with heart rate monitors and body pose sensors that can identify symptoms in the context of postprandial phases or while the user is lying down.

Musculoskeletal Manifestations

Muscle weakness and joint pain have been reported among PASC patients, and the signs associated with them may be well quantified using heart and respiratory rate monitors or body pose sensors, particularly in cases for which there is a decreased range of motion. Pain that affects a patient's ability to perform activities of daily living may manifest as behavior changes and can be assessed with activity monitors. Electrodermal conductance has been shown to increase in the setting of pain, so monitoring skin conductance could also provide insight into the pathophysiology of PASC [58,59]. Some work has also been done to correlate facial expressions with quantitative measures of pain [60].

General Well-being Manifestations

Fatigue is an extremely common manifestation of PASC and is one of the most common presenting symptoms. It has been suggested that electroencephalogram (EEG) measurements can also be used to detect symptoms of fatigue, but until EEG technologies can be used discreetly by users, their use will likely be limited to controlled clinical and research settings [61]. In lieu of using EEG measurements, activity type and level can easily be measured to quantify fatigue symptoms, and while not specific for fatigue, they can be used by knowledgeable care providers in combination with other clinical measures to better understand a patient's presenting signs. Such activity measurements can be made using either wearable or contactless solutions as presented in the previous sections [62,63].

Other Reported Symptoms

Other reported symptoms of PASC that are recognized by the CDC that have not yet been presented in this viewpoint include fever, rash, and changes in menstrual cycles. Signs of fever can most directly be quantified using sensors such as infrared thermographic cameras [64]. New-onset rashes may be best temporally tagged with activity monitoring that identifies new or worsening instances of itching. Wearable sensor patches have been developed that can sense the act of itching and can be utilized in the setting of PASC [65]. Alternatively, cameras trained to recognize users may be able to identify new visible lesions, though such technology is currently far from being commercially or medically viable and alternative non-camera sensing is preferred as users have demonstrated concerns of poor privacy and discomfort with camera-based sensors [66]. Changes in menstrual cycles may be tracked, in part, using heart rate and/or temperature sensors during periods of sleep [67,68].

Sensor Technologies

Overview

Several continuous sensing technologies are currently available for consumer use, and more are in research or development. Currently, the majority of sensors fall under the category of wearables. These products, like the Apple Watch, the Fitbit Sense, or the Samsung Galaxy Watch3, have the benefit of being portable, though users must remember to constantly charge and wear the products to collect data. These pose physical and cognitive burdens to patients, and achieving long-term compliance can be difficult, especially for those with impaired dexterity or cognition. Aside from wearable sensors, there is growing interest in engineering circles to develop contactless sensors. Such sensors usually obtain measures without touching the human body and do not require any physical or cognitive efforts from the user. Figure 1 illustrates a prototype designed at Stony Brook University using off-the-shelf components, in particular an ultra-wideband radio that transmits extremely short pulses that reflect off the human body [69]. The chest movements from respirations and heart beats cause changes in the reflected signal. After sophisticated signal processing, the

heart and respiration rates can be extracted (shown on the computer screen). Such a device can derive pulse data that are comparable to those of an FDA-approved fingertip oximeter (Masimo MightySat), whose data are shown on an iPad. The depth camera, a special device that generates coarse-grained 3-dimensional body joint locations but not detailed regular RGB images, can be used to recognize the body pose and thus activities, without revealing detailed visual images that cause privacy concerns. The data would be analyzed in a secure computing environment so as to enhance privacy protection. Nevertheless, privacy considerations represent a potential barrier to widespread sensor adoption, and care should be taken to not only protect patient privacy with utmost care but also convey to potential users the protective measures taken in developing sensor technologies.

Figure 1. Illustration of a contactless sensor using an ultra-wideband (UWB) radio system on chip (SoC) and a depth camera. The radio waves reflect off the chest and become modulated by respiration and heartbeat movement and are received and processed in order to extract heart rate (HR) and respiration rate (RR). Such contactless sensors can produce results similar to fingertip oximeter pulse readings. This sensor is a prototype that we envision will be housed in a small-form customized hardware package. RGB: red, green, blue; TX: transmitting antenna; RX: receiving antenna.

When properly packaged (eg, putting the radio and depth camera in one case), such contactless sensors can be mounted on the wall and left to collect data continuously while the user simply conducts his or her daily routines as usual. There exist other contactless sensors, including the sleep monitoring functionality of the SleepScore Max and the recently released Google Nest Hub. They are only now beginning to enter the consumer market [70,71]. They offer substantial benefits, particularly among older individuals who, as they age, spend increasing amounts of time in the home; those with progressing chronic diseases; and individuals with cognitive impairment or decline.

Wearable Sensors

Wearable sensors include anything that can collect data about a user and requires a physical proximity (usually by touch) between the user and device to operate. Wearable sensors are a widely saturated component of the current market. Often, they are sought for reasons other than their sensor functionality. Smartwatches are a particularly demonstrative example of this. Although features of the Apple Watch series 6 include pulse oximetry, actigraphy, photoplethysmography, electrocardiography, pedometry, and altimetry, its mass appeal to many stems from its ability to connect with other devices, allowing for a portable means to listen to music, access text messages, take phone calls, and check the time. Such benefits make them widely sought and already frequently used. Other continuously or regularly monitoring wearable devices include chest-strap heart rate monitors, wearable blood oxygen saturation (SpO_2) trackers, continuous positive airway pressure (CPAP) machines, smart clothing, smart hearing aids, smart headphones, and smart glasses.

Wearable sensors are widely versatile, in part due to their portability. The tradeoff, however, is that wearable sensors need to be periodically charged and worn. For some, this may be habitual, but for others, adherence is a significant consideration that may make data collection less effective. In the context of PASC, when continuous, contextualized data collections are integral to a holistic approach to care, complementary sensor capabilities should be considered for these patients. For some users, contactless sensors may offer these complementary capabilities.

Contactless Sensors

Contactless sensors include ambient sensors that can be strategically implemented throughout an individuals' physical space. Such sensors can measure light, motion, vibration, pressure, and electromagnetic echoes from a distance and can be used to monitor vital signs, activity patterns, sleep quality, and activities of daily living. Newer prototypes can detect movement at such granularity as to resolve minute chest-wall motions from breaths and heartbeats [69,72,73]. With the widespread implementation of nonwearable sensors, there would be less of an onus on patients to ensure that their devices are fully charged and that their sensors are being worn consistently and correctly. Rather, contactless sensors provide a framework that makes passive monitoring possible without any cooperative efforts from the user. Such a system would allow physicians and other care providers to access continuously collected data on activity and vital signs from patients in a home or home-like environment provided patient consent to that data collection. This stands in juxtaposition to wearable sensors with which patients must actively cooperate (ie, charging, wearing) prior

to data being acquired. These features of contactless sensing have made them a natural choice in monitoring ailments associated with aging, often used instead of or in conjunction with the better-established wearable sensors [74].

The added benefit of contactless sensors is the fact that some of them can be used to track multiple subjects. This cost-saving measure could be particularly well utilized in the context of PASC as acute COVID was, and continues to be, rapidly disseminated in household and nursing homes clusters, where a single sensor could be used by multiple users [75,76]. The use of contactless sensors, then, would minimize the resources necessary to effectively monitor and track PASC within high-exposure households. Nevertheless, the use of individual contactless sensors may have limited sensing range or angle, thus not offering full coverage without the installation of multiple sensors, motivating the complementary need for wearable sensors.

Discussion

Currently, health care professionals assign a diagnosis of PASC when a patient who had previously been diagnosed with COVID-19 or who had a strong clinical suspicion of having had COVID-19 has ongoing or new symptoms of otherwise unexplainable etiology that arose after the acute COVID-19 phase. Such patients are now commonly managed in newly designated PASC clinics offering periodic professional consultation and symptom management for this heretofore poorly understood disease [77]. Before enrollment in a PASC clinic, few in-hospital tests can be conducted to further support the diagnosis of PASC. These diagnostics, including basic laboratory testing, serological screening, targeted history and physical taking, and stress testing can be viewed on the CDC website [78]. These tests are limited in their ability to quantify symptoms in waxing and waning cycles of symptom intensity. In addition to this general and unspecific approach to diagnosis, treatment for PASC, when available, relies on reactive management of symptoms. These challenges in monitoring and managing symptoms can, in part, be addressed with the widespread adoption of sensors; sensors can facilitate continuously monitoring a patient by multimodality data collection. This monitoring, in concert with analysis and interpretation of existing tools that have diagnostic power, can provide a more holistic picture of PASC and its pathophysiological progression.

Continuous sensing devices like smartwatches with even basic measurement capabilities would be instrumental in clarifying for the medical community the temporal development of PASC. Furthermore, it would be useful in differentiating symptom clusters of PASC as has recently been observed [79]. Such devices, where feasible, should be provided to patients at initial presentation with COVID-19. For patients who already have and use continuous sensing devices, a previously established baseline of vital signs would help track symptom abatement. Such deviation-from-baseline tracking has already been used to predict acute COVID-19 before the onset of clinical symptoms and has begun to be used in the context of PASC [79,80].

With regards to symptom monitoring and tracking, there is substantial utility in the use of sensors to derive context-specific symptom management. Many symptoms associated with PASC are exacerbated, and some are alleviated by performing certain tasks. Many patients complain about having difficulties performing activities of daily living (ie, climbing stairs, standing). Having quantitative time-linked, or in the case of some contactless sensors, directly activity-linked, vital sign metrics would be helpful in both clinical management and research. With context-specific data, symptom management can become pre-emptive rather than reactive. Furthermore, with quantitative measures of symptoms, occupational and physical therapists can better address difficulties in performing specific daily tasks. Additionally, while many of the signs associated with each of the above common symptoms are nonspecific, thorough cataloging of disease sign presentations will be extremely impactful in shaping management and understanding.

Ultimately, the utility of continuous sensing devices in the context of PASC has both clinical and academic/research utility. Such data collection will not only provide better-targeted care to the individual patients with persisting COVID symptoms but will also contribute to the medical understanding of how PASC manifests. Such information is critical in the data-poor arena of PASC. The use of such sensors in the monitoring of chronic diseases is not a new concept, and existing projects have been met with success. Programs have been implemented to both track and facilitate activity in individuals with chronic disease and to monitor elderly populations to facilitate aging in place [18,81-85]. These patient populations have benefitted largely from the volume of data collected to improve the quality of life for patients and to better describe malady in general. Given the global burden of the COVID-19 pandemic, a similar strategy to those already being implemented should be initiated to determine how best to allocate resources to individuals with PASC symptoms. Such investment in monitoring, especially monitoring by passive means due to the lack of extra user efforts, will take some of the burden off medical professionals and health care resources by streamlining care, allowing physicians a more holistic understanding of their patients' medical presentations, and providing them a foundation on which to prioritize management strategies. The net result of such widespread adoption of innovative sensor technologies would be to facilitate the diminution of some of the long-term systemic drains on the medical system due to the pandemic.

There is also a social component to PASC that can be addressed with sensors. PASC has been associated with the stigma of being psychosomatic rather than physical in certain segments of the population [86]. Furthermore, many do not recognize their symptoms as being significant enough to warrant seeing a doctor or may fail to connect their symptoms with PASC [87]. With continuous sensing, the timeline associated with the onset of symptoms could be retroactively assessed by a clinician, allowing for streamlined medical care despite delays in initially seeking that care. Though the use of wearable and contactless sensors for monitoring health is still limited in its adoption, such technology offers patients and providers significant benefits in the identification and management of PASC.

Although continuous sensing promises a substantial benefit to the health care system in the context of PASC, it is worth considering the potential challenges of a widespread continuous sensing paradigm. First, as many of these technologies are still in development and their use outside of a theoretical or research framework has so far been limited, few have received FDA approval for use in clinical management. Most will likely not receive approval within any practical timeline. As such, sensor data should be used at providers' discretion, and data should be used to augment clinical practice and not replace it. For technologies that have gone through rigorous studies, their reliability and validity could be factored into their role in clinical care and should be considered on a case-by-case basis. Many such technologies have been demonstrated to be relatively accurate when compared with a medical gold standard [69,88,89]. However, it is not advised that non-FDA-approved devices be used for the diagnosis or management of any disease without substantial clinical suspicion to corroborate. We therefore envision the role of sensors to be adjunctive to the use of existing FDA-approved technologies. We stress that any abnormal or concerning finding obtained from non-FDA-approved technologies should trigger subsequent follow-up. Second, it is important to consider how the influx of additional information will be handled by health care professionals. Sensors can quantify physiologic signs. Associating those signs with specific symptoms is important in developing a clinical picture but requires careful interpretation of the data in conjunction with clinical judgment. Sensor data do not represent a replacement for clinical intuition. Instead, sensors promise a modality by which suspicions may be corroborated and subsequent confirmatory workup can be initiated. In instances where sensor data fail to corroborate a hypothesized pathophysiologic course, it may be impossible to determine whether the sensor(s) or the algorithms that interpret the data lack sufficient sensitivity or there is a true lack of symptom(s). The inability to resolve the cause of these discrepant measurements may lead to patient anxiety. It is, therefore, important for physicians to temper these expectations in patients and to understand the limitations inherent to any given sensing system. It should be stressed to the patient that continuous sensing can be effective at offering additional evidence correlated with suspected signs and symptoms, but it is inherently limited due to the physics of sensing hardware (ie, noises, disturbances) and the abilities of algorithms (ie, fidelity achievable interpreting signals), thus should only be used to aid but never to replace clinical measurements and diagnosis. Sensors should never serve as the sole foundation of a clinical picture or clinical management without evaluating all clinical data available.

Finally, it is also important to consider barriers to widespread sensor adoption from the perspective of users. Since sensors are effective at collecting vital sign data, there is an inherent privacy concern that may prevent their use. Care must be taken to implement Health Insurance Portability and Accountability Act (HIPAA)–compliant protocols for the encryption and protection of patient information. As part of these protocols, patients should be empowered to withhold sensor data from anyone, including their health care providers, should they so choose. The responsibility of ensuring that a given sensor sufficiently prioritizes privacy will likely fall on both the manufacturer and the individual represented at the point of acquiring the sensor (ie, retail salesperson, physician). These parties must have a practical understanding of privacy measures taken to accurately convey to the potential user the risks and benefits of using the product. Furthermore, there is a challenge of equity. The cost of purchasing sensors, either by institutions or users directly, may be prohibitive. As such, providing access to these sensor technologies in regions of low socioeconomic status may be challenging. Given the demonstrated benefits of sensors, we recommend that physicians or health care systems be empowered to provide or lend sensor technologies to those patients who stand to benefit from them, particularly those patients who cannot afford sensors themselves. Short of hospital and clinic-sponsored dissemination of sensor technologies, on a federal level, allocating resources to subsidize the purchase of vital sensors for those who cannot afford them would facilitate equity in sensor utilization. In regions that lack internet connectivity or reliable access to clinics or hospitals, the lack of existing infrastructure may make sensor utilization less effective. Providing the infrastructure necessary to accommodate these sensors is itself a public health challenge and has garnered widespread attention. It is of vital importance that reliable internet and stable access to care are ensured for all. Though such approaches to addressing health care infrastructure deficits fall beyond the scope of this paper, it is of our opinion that such basic functions should be afforded to all.

Despite these challenges, with trained health care professionals making decisions with the help of continuous sensing data, management plans can be more effectively designed and implemented, and a more comprehensive academic understanding of PASC can be achieved. Such an approach to disease tracking has been used in the context of acute COVID-19, with particular emphasis on prioritizing sensors that track respiratory rate, pulse oximetry, heart rate, heart rhythm, and blood pressure [90]. With the widespread recognition of the utility of sensors in disease tracking generally, and in COVID specifically, it is imperative that there is continued and renewed investment in developing continuous sensing technologies with relatively high home deployment feasibility and more universal adoption of such technologies between clinical visits. Particularly as segments of medical practice have adopted measures to accommodate telemedicine, the importance of a passive means of garnering quantitative patient data with real-time evolution collected outside of medical establishments is difficult to overstate [91]. Furthermore, care must be made to develop rigorous analytics software to interpret raw sensor output data into actionable insights impacting clinical decisions. Laying the groundwork for such a technologically motivated health care infrastructure would provide medical professionals a plethora of data to augment existing clinical measures for targeted, individualized health care decisions with their patients.

Although PASC is, as yet, a poorly understood disease, a symptom-based approach to tracking its pathophysiology may inform management decisions. Many of the symptoms associated with PASC are relatively well-studied, and while curative treatments may not be universally applicable to target

each symptom, palliative treatments that target specific symptoms may be available. With a rigorous quantification of symptoms, clinicians can strategize treatment plans, thereby mitigating wasteful hospital spending on ineffective treatments [92]. Given their variable but often substantial cost, sensor technologies, particularly novel technologies, will need to be paid for by, subsidized by, or rented out from federal, state, or private entities in order to maximize access and equitable use. However, the benefit of tracking therapeutic effects represents a huge potential for net savings. Although this paper focuses on the application of continuous unobtrusive sensors in the management of PASC, broader applications exist for the tracking of chronic diseases generally, with estimates of spending on inadequate treatment of chronic disease as high as US $2.5 billion annually for rheumatoid arthritis patients alone, representing a substantial outlet of wasted spending that these sensors can begin to address [93]. Furthermore, patients at risk of particularly debilitating disease can be identified early in their recovery and directed to appropriate resources (ie, pharmacotherapy, physical therapy, pulmonary therapy) to preempt deterioration, thereby enhancing the efficacy of intervention with early initiation [92]. In such instances where treatments are initiated, unobtrusive or contactless sensors are well-equipped to track physiological changes associated with the intervention, and thus widespread utilization of such sensors may serve as a basis for the iterative streamlining of health interventions. Interventions that yield ineffective or deleterious effects will be easily identified and removed or replaced until optimal treatment regimens can be reached.

Such an approach promises high potential for significant improvement in the quality of life for PASC patients and a more efficient utilization of health care spending. Similar symptom-tied approaches to continuous sensing will likely be of significant clinical utility in the management of other chronic diseases as well. The NIH recognizes this potential. In their recently released NIH-Wide Strategic Plan for fiscal years 2021-2025, they acknowledge the importance of being able to track patient signs across time, outside of a clinical setting [94]. The document also emphasizes the utility of “sensors that can provide continuous feedback,” particularly because of their ability to “detect underlying signs of illness and response to intervention, including medications and lifestyle changes faster than conventional methods.” Such a formal recognition of the importance of continuous sensing will be invaluable in reinforcing the academic credibility of such further study.

The COVID-19 pandemic has forced a renewed awareness of health care delivery systems and the technologies that define them. It is imperative, then, that we in the medical and engineering communities engage with the public discourse of the present moment and leverage the development, adoption, and interpretation of technologies that streamline care and yield promising health outcomes. We consider continuous sensing to be an essential component of this modern health care system. Such a goal of streamlined and optimal care necessitates interprofessional collaboration and community engagement. It is our hope that this viewpoint provides suggestions and inspiration for future interdisciplinary collaborations that facilitate the innovation of sensor technologies and the infrastructure that supports them.

Acknowledgments

The authors would like to thank Dr Norman Edelman, MD for his insight and expertise. This work was supported in part by the National Science Foundation (NSF) grants #2028952 and #2119299.

Conflicts of Interest

None declared.

References

1. COVID Data Tracker. Centers for Disease Control and Prevention. URL: https://covid.cdc.gov/covid-data-tracker/#datatracker-home [accessed 2022-01-15]
2. COVID-19 Vaccines. Food & Drug Administration. URL: https://www.fda.gov/emergency-preparedness-and-response/coronavirus-disease-2019-covid-19/covid-19-vaccines [accessed 2022-01-15]
3. Post-COVID-19 Conditions. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects.html#:~:text=Tiredness%20or%20fatigue.%20Difficulty,or%20shortness%20of%20breath [accessed 2022-01-15]
4. Taquet M, Dercon Q, Luciano S, Geddes JR, Husain M, Harrison PJ. Incidence, co-occurrence, and evolution of long-COVID features: A 6-month retrospective cohort study of 273,618 survivors of COVID-19. PLoS Med 2021 Sep;18(9):e1003773 [FREE Full text] [doi: 10.1371/journal.pmed.1003773] [Medline: 34582441]
5. Callard F, Perego E. How and why patients made Long Covid. Soc Sci Med 2021 Jan;268:113426 [FREE Full text] [doi: 10.1016/j.socscimed.2020.113426] [Medline: 33199035]
6. When COVID-19 Symptoms Linger. National Institutes of Health. URL: https://covid19.nih.gov/news-and-stories/when-COVID-19-symptoms-linger [accessed 2022-01-15]
7. Lambert N, Survivor Corps. COVID-19 "Long Hauler" Symptoms Survey Report. Indiana University School of Medicine. 2020. URL: https://scholarworks.iupui.edu/bitstream/handle/1805/25685/Lambert2020COVID-19.pdf [accessed 2022-01-15]

8. Twomey R, DeMars J, Franklin K, Culos-Reed S, Weatherald J, Wrightson J. Chronic fatigue and post-exertional malaise in people living with long COVID. medRxiv. 2021. URL: https://www.medrxiv.org/content/10.1101/2021.06.11.21258564v1 [accessed 2022-01-15]
9. Piasecki TM, Hufford MR, Solhan M, Trull TJ. Assessing clients in their natural environments with electronic diaries: rationale, benefits, limitations, and barriers. Psychol Assess 2007 Mar;19(1):25-43. [doi: 10.1037/1040-3590.19.1.25] [Medline: 17371121]
10. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient compliance with paper and electronic diaries. Control Clin Trials 2003 Apr;24(2):182-199. [doi: 10.1016/s0197-2456(02)00320-3] [Medline: 12689739]
11. Arga M, Sahbaz H, Bakirtas A, Turktas I, Demirsoy MS. Does self-monitoring by means of symptom diaries improve asthma control in children? J Asthma 2014 Apr;51(3):299-305. [doi: 10.3109/02770903.2013.867974] [Medline: 24274828]
12. Burton C, Weller D, Sharpe M. Are electronic diaries useful for symptoms research? A systematic review. J Psychosom Res 2007 May;62(5):553-561. [doi: 10.1016/j.jpsychores.2006.12.022] [Medline: 17467410]
13. Green AS, Rafaeli E, Bolger N, Shrout PE, Reis HT. Paper or plastic? Data equivalence in paper and electronic diaries. Psychological Methods 2006 Mar;11(1):87-105. [doi: 10.1037/1082-989x.11.1.87]
14. Dani M, Dirksen A, Taraborrelli P, Torocastro M, Panagopoulos D, Sutton R, et al. Autonomic dysfunction in 'long COVID': rationale, physiology and management strategies. Clin Med (Lond) 2021 Jan;21(1):e63-e67 [FREE Full text] [doi: 10.7861/clinmed.2020-0896] [Medline: 33243837]
15. Subbiah R, Gula LJ, Klein GJ, Skanes AC, Yee R, Krahn AD. Syncope: review of monitoring modalities. Curr Cardiol Rev 2008 Feb;4(1):41-48 [FREE Full text] [doi: 10.2174/157340308783565447] [Medline: 19924276]
16. Juen J, Cheng Q, Prieto-Centurion V, Krishnan JA, Schatz B. Health monitors for chronic disease by gait analysis with mobile phones. Telemed J E Health 2014 Nov;20(11):1035-1041 [FREE Full text] [doi: 10.1089/tmj.2014.0025] [Medline: 24694291]
17. Dieffenderfer J, Goodell H, Mills S, McKnight M, Yao S, Lin F, et al. Low-power wearable systems for continuous monitoring of environment and health for chronic respiratory disease. IEEE J Biomed Health Inform 2016 Sep;20(5):1251-1264 [FREE Full text] [doi: 10.1109/JBHI.2016.2573286] [Medline: 27249840]
18. Chowdhury A, Bhattacharya S, Ghose A, Krishnan B. Early detection of mild cognitive impairment using pervasive sensing. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:5456-5459. [doi: 10.1109/EMBC.2019.8856435] [Medline: 31947090]
19. Atallah L, Lo B, Yang G. Can pervasive sensing address current challenges in global healthcare? J Epidemiol Glob Health 2012 Mar;2(1):1-13 [FREE Full text] [doi: 10.1016/j.jegh.2011.11.005] [Medline: 23856393]
20. Chan PY, McNeil J, Nguyen T, Ryan N, Hopper I. Novel wearable and contactless monitoring devices to identify deteriorating patients in the clinical setting: a systematic review protocol. Syst Rev 2020 May 06;9(1):104 [FREE Full text] [doi: 10.1186/s13643-020-01370-1] [Medline: 32375867]
21. COVID-19 rapid guideline: managing the long-term effects of COVID-19. National Institute for Health and Care Excellence. 2021 Nov 11. URL: https://www.nice.org.uk/guidance/ng188 [accessed 2022-01-15]
22. A clinical case definition of post COVID-19 condition by a Delphi consensus. World Health Organization. 2021 Oct 06. URL: https://www.who.int/publications/i/item/WHO-2019-nCoV-Post_COVID-19_condition-Clinical_case_definition-2021.1 [accessed 2022-01-15]
23. Post-COVID Conditions. Centers for Disease Control and Prevention. 2021 Sep 16. URL: https://www.cdc.gov/coronavirus/2019-ncov/long-term-effects/index.html [accessed 2022-01-15]
24. Greenhalgh T, Knight M, A'Court C, Buxton M, Husain L. Management of post-acute covid-19 in primary care. BMJ 2020 Aug 11;370:m3026. [doi: 10.1136/bmj.m3026] [Medline: 32784198]
25. Al-Aly Z, Xie Y, Bowe B. High-dimensional characterization of post-acute sequelae of COVID-19. Nature 2021 Jun;594(7862):259-264. [doi: 10.1038/s41586-021-03553-9] [Medline: 33887749]
26. Hageman JR. Long COVID-19 or post-acute sequelae of SARS-CoV-2 infection in children, adolescents, and young adults. Pediatr Ann 2021 Jun;50(6):e232-e233 [FREE Full text] [doi: 10.3928/19382359-20210519-02] [Medline: 34115558]
27. Nalbandian A, Sehgal K, Gupta A, Madhavan MV, McGroder C, Stevens JS, et al. Post-acute COVID-19 syndrome. Nat Med 2021 Apr;27(4):601-615. [doi: 10.1038/s41591-021-01283-z] [Medline: 33753937]
28. Peghin M, Palese A, Venturini M, De Martino M, Gerussi V, Graziano E, et al. Post-COVID-19 symptoms 6 months after acute infection among hospitalized and non-hospitalized patients. Clin Microbiol Infect 2021 Oct;27(10):1507-1513 [FREE Full text] [doi: 10.1016/j.cmi.2021.05.033] [Medline: 34111579]
29. Hirschtick JL, Titus AR, Slocum E, Power LE, Hirschtick RE, Elliott MR, et al. Population-based estimates of post-acute sequelae of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection (PASC) prevalence and characteristics. Clin Infect Dis 2021 Dec 06;73(11):2055-2064 [FREE Full text] [doi: 10.1093/cid/ciab408] [Medline: 34007978]
30. Cabrera Martimbianco AL, Pacheco RL, Bagattini, Riera R. Frequency, signs and symptoms, and criteria adopted for long COVID-19: A systematic review. Int J Clin Pract 2021 Oct 02;75(10):e14357 [FREE Full text] [doi: 10.1111/ijcp.14357] [Medline: 33977626]
31. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More than 50 long-term effects of COVID-19: a systematic review and meta-analysis. Sci Rep 2021 Aug 09;11(1):16144 [FREE Full text] [doi: 10.1038/s41598-021-95565-8] [Medline: 34373540]

32. Nasserie T, Hittle M, Goodman SN. Assessment of the frequency and variety of persistent symptoms among patients with COVID-19: a systematic review. JAMA Netw Open 2021 May 03;4(5):e2111417 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.11417] [Medline: 34037731]
33. Polsky MB, Moraveji N. Post-acute care management of a patient with COVID-19 using remote cardiorespiratory monitoring. Respir Med Case Rep 2021;33:101436 [FREE Full text] [doi: 10.1016/j.rmcr.2021.101436] [Medline: 34150489]
34. Chu M, Nguyen T, Pandey V, Zhou Y, Pham HN, Bar-Yoseph R, et al. Respiration rate and volume measurements using wearable strain sensors. NPJ Digit Med 2019;2:8 [FREE Full text] [doi: 10.1038/s41746-019-0083-3] [Medline: 31304358]
35. Torres-Castro R, Vasconcello-Castillo L, Alsina-Restoy X, Solis-Navarro L, Burgos F, Puppo H, et al. Respiratory function in patients post-infection by COVID-19: a systematic review and meta-analysis. Pulmonology 2021;27(4):328-337 [FREE Full text] [doi: 10.1016/j.pulmoe.2020.10.013] [Medline: 33262076]
36. Núñez-Cortés R, Rivera-Lillo G, Arias-Campoverde M, Soto-García D, García-Palomera R, Torres-Castro R. Use of sit-to-stand test to assess the physical capacity and exertional desaturation in patients post COVID-19. Chron Respir Dis 2021;18:1479973121999205 [FREE Full text] [doi: 10.1177/1479973121999205] [Medline: 33645261]
37. Goodacre S, Thomas B, Lee E, Sutton L, Loban A, Waterhouse S, et al. Post-exertion oxygen saturation as a prognostic factor for adverse outcome in patients attending the emergency department with suspected COVID-19: a substudy of the PRIEST observational cohort study. Emerg Med J 2021 Feb;38(2):88-93 [FREE Full text] [doi: 10.1136/emermed-2020-210528] [Medline: 33273040]
38. Kalin A, Javid B, Knight M, Inada-Kim M, Greenhalgh T. Direct and indirect evidence of efficacy and safety of rapid exercise tests for exertional desaturation in Covid-19: a rapid systematic review. Syst Rev 2021 Mar 16;10(1):77 [FREE Full text] [doi: 10.1186/s13643-021-01620-w] [Medline: 33726854]
39. Crooks MG, den Brinker A, Hayman Y, Williamson JD, Innes A, Wright CE, et al. Continuous cough monitoring using ambient sound recording during convalescence from a COPD exacerbation. Lung 2017 Jun;195(3):289-294 [FREE Full text] [doi: 10.1007/s00408-017-9996-2] [Medline: 28353117]
40. Marston HR, Hadley R, Banks D, Duro MDCM. Mobile self-monitoring ECG devices to diagnose arrhythmia that coincide with palpitations: a scoping review. Healthcare (Basel) 2019 Aug 16;7(3):1 [FREE Full text] [doi: 10.3390/healthcare7030096] [Medline: 31426383]
41. Zimetbaum P, Goldman A. Ambulatory arrhythmia monitoring: choosing the right device. Circulation 2010 Oct 19;122(16):1629-1636. [doi: 10.1161/CIRCULATIONAHA.109.925610] [Medline: 20956237]
42. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, et al. Smart wearable devices in cardiovascular care: where we are and how to move forward. Nat Rev Cardiol 2021 Aug 04;18(8):581-599 [FREE Full text] [doi: 10.1038/s41569-021-00522-7] [Medline: 33664502]
43. Johansson M, Ståhlberg M, Runold M, Nygren-Bonnier M, Nilsson J, Olshansky B, et al. Long-haul post-COVID-19 symptoms presenting as a variant of postural orthostatic tachycardia syndrome: the Swedish experience. JACC Case Rep 2021 Apr;3(4):573-580 [FREE Full text] [doi: 10.1016/j.jaccas.2021.01.009] [Medline: 33723532]
44. Zaffaroni A, Coffey S, Dodd S, Kilroy H, Lyon G, O'Rourke D, et al. Sleep staging monitoring based on sonar smartphone technology. Annu Int Conf IEEE Eng Med Biol Soc 2019 Jul;2019:2230-2233. [doi: 10.1109/EMBC.2019.8857033] [Medline: 31946344]
45. Rawtaer I, Mahendran R, Kua EH, Tan HP, Tan HX, Lee T, et al. Early detection of mild cognitive impairment with in-home sensors to monitor behavior patterns in community-dwelling senior citizens in Singapore: cross-sectional feasibility study. J Med Internet Res 2020 May 05;22(5):e16854 [FREE Full text] [doi: 10.2196/16854] [Medline: 32369031]
46. Li X, He Y, Jing X. A survey of deep learning-based human activity recognition in radar. Remote Sensing 2019 May 06;11(9):1068. [doi: 10.3390/rs11091068]
47. He Y, Li Y. Physical activity recognition utilizing the built-in kinematic sensors of a smartphone. International Journal of Distributed Sensor Networks 2013 Apr 10;9(4):481580. [doi: 10.1155/2013/481580]
48. Budson AE. What is COVID-19 brain fog — and how can you clear it? Harvard Health Publishing. 2021 Mar 08. URL: https://www.health.harvard.edu/blog/what-is-covid-19-brain-fog-and-how-can-you-clear-it-2021030822076 [accessed 2022-01-15]
49. Wong AA, Read SJ. Early changes in physiological variables after stroke. Ann Indian Acad Neurol 2008 Oct;11(4):207-220 [FREE Full text] [doi: 10.4103/0972-2327.44555] [Medline: 19893676]
50. Sathish T, Anton MC, Sivakumar T. New-onset diabetes in "long COVID". J Diabetes 2021 Aug;13(8):693-694 [FREE Full text] [doi: 10.1111/1753-0407.13187] [Medline: 33893710]
51. Aminian A, Bena J, Pantalone KM, Burguera B. Association of obesity with postacute sequelae of COVID-19. Diabetes Obes Metab 2021 Sep;23(9):2183-2188 [FREE Full text] [doi: 10.1111/dom.14454] [Medline: 34060194]
52. Rubino F, Amiel SA, Zimmet P, Alberti G, Bornstein S, Eckel RH, et al. New-onset diabetes in Covid-19. N Engl J Med 2020 Aug 20;383(8):789-790 [FREE Full text] [doi: 10.1056/NEJMc2018688] [Medline: 32530585]
53. Lorenz KA, Yeshurun S, Aziz R, Ortiz-Delatorre J, Bagley JR, Mor M, et al. A handheld metabolic device (Lumen) to measure fuel utilization in healthy young adults: device validation study. Interact J Med Res 2021 May 17;10(2):e25371 [FREE Full text] [doi: 10.2196/25371] [Medline: 33870899]

54. Wang C, Li X, Hu H, Zhang L, Huang Z, Lin M, et al. Monitoring of the central blood pressure waveform via a conformal ultrasonic device. Nat Biomed Eng 2018 Sep;2(9):687-695 [FREE Full text] [doi: 10.1038/s41551-018-0287-x] [Medline: 30906648]
55. Chakraborty S, Mandal J, Yang T, Cheng X, Yeo J, McCarthy CG, et al. Metabolites and hypertension: insights into hypertension as a metabolic disorder: 2019 Harriet Dustan Award. Hypertension 2020 Jun;75(6):1386-1396 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.120.13896] [Medline: 32336227]
56. van Helmond N, Freeman CG, Hahnen C, Haldar N, Hamati JN, Bard DM, et al. The accuracy of blood pressure measurement by a smartwatch and a portable health device. Hosp Pract (1995) 2019 Oct;47(4):211-215. [doi: 10.1080/21548331.2019.1656991] [Medline: 31423912]
57. Riaz MS, Atreja A. Personalized technologies in chronic gastrointestinal disorders: self-monitoring and remote sensor technologies. Clin Gastroenterol Hepatol 2016 Dec;14(12):1697-1705 [FREE Full text] [doi: 10.1016/j.cgh.2016.05.009] [Medline: 27189911]
58. Naranjo-Hernández D, Reina-Tosina J, Roa LM. Sensor technologies to manage the physiological traits of chronic pain: a review. Sensors (Basel) 2020 Jan 08;20(2):1 [FREE Full text] [doi: 10.3390/s20020365] [Medline: 31936420]
59. Rodríguez I, Herskovic V, Gerea C, Fuentes C, Rossel PO, Marques M, et al. Understanding monitoring technologies for adults with pain: systematic literature review. J Med Internet Res 2017 Oct 27;19(10):e364 [FREE Full text] [doi: 10.2196/jmir.7279] [Medline: 29079550]
60. Bargshady G, Soar J, Zhou X, Deo R, Whittaker F, Wang H. A Joint Deep Neural Network Model for Pain Recognition from Face. 2019 Presented at: IEEE 4th International Conference on Computer and Communication Systems (ICCCS); February 23-25, 2019; Singapore. [doi: 10.1109/ccoms.2019.8821779]
61. Kartsch VJ, Benatti S, Schiavone PD, Rossi D, Benini L. A sensor fusion approach for drowsiness detection in wearable ultra-low-power systems. Information Fusion 2018 Sep;43:66-76. [doi: 10.1016/j.inffus.2017.11.005]
62. Gholami M, Napier C, Patiño AG, Cuthbert TJ, Menon C. Fatigue monitoring in running using flexible textile wearable sensors. Sensors (Basel) 2020 Sep 29;20(19):1 [FREE Full text] [doi: 10.3390/s20195573] [Medline: 33003316]
63. Luo H, Lee P, Clay I, Jaggi M, De Luca V. Assessment of fatigue using wearable sensors: a pilot study. Digit Biomark 2020;4(Suppl 1):59-72 [FREE Full text] [doi: 10.1159/000512166] [Medline: 33442581]
64. Abuzairi T, Imaniati Sumantri N, Irfan A, Maulana Mohamad R. Infrared thermometer on the wall (iThermowall): An open source and 3-D print infrared thermometer for fever screening. HardwareX 2021 Apr;9:e00168 [FREE Full text] [doi: 10.1016/j.ohx.2020.e00168] [Medline: 33392421]
65. Chun KS, Kang YJ, Lee JY, Nguyen M, Lee B, Lee R, et al. A skin-conformable wireless sensor to objectively quantify symptoms of pruritus. Sci Adv 2021 Apr;7(18):1 [FREE Full text] [doi: 10.1126/sciadv.abf9405] [Medline: 33931455]
66. Caine K, Šabanović S, Carter M. The effect of monitoring by cameras and robots on the privacy enhancing behaviors of older adults. 2012 Presented at: 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI); March 5-8, 2012; Boston, MA. [doi: 10.1145/2157689.2157807]
67. Shilaih M, Clerck VD, Falco L, Kübler F, Leeners B. Pulse rate measurement during sleep using wearable sensors, and its correlation with the menstrual cycle phases, a prospective observational study. Sci Rep 2017 May 02;7(1):1294 [FREE Full text] [doi: 10.1038/s41598-017-01433-9] [Medline: 28465583]
68. Hessemer V, Brück K. Influence of menstrual cycle on shivering, skin blood flow, and sweating responses measured at night. J Appl Physiol (1985) 1985 Dec;59(6):1902-1910. [doi: 10.1152/jappl.1985.59.6.1902] [Medline: 4077797]
69. Xie Z, Zhou B, Cheng X, Schoenfeld E, Ye F. VitalHub: Robust, Non-Touch Multi-User Vital Signs Monitoring using Depth Camera-Aided UWB. In: IEEE 9th International Conference on Healthcare Informatics (ICHI). 2021 Presented at: IEEE 9th International Conference on Healthcare Informatics (ICHI); August 9-12, 2021; Victoria, BC, Canada. [doi: 10.1109/ichi52183.2021.00056]
70. Dixon M, Schneider L, Yu J, Hsu J, Pathak A, Shin D, et al. Sleep-wake Detection With a Contactless, Bedside Radar Sleep Sensing System. Google. 2021. URL: https://research.google/pubs/pub50198/ [accessed 2022-01-15]
71. Chinoy E, Cuellar J, Huwa K, Jameson J, Watson C, Bessman S, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep 2021 May 14;44(5):1 [FREE Full text] [doi: 10.1093/sleep/zsaa291] [Medline: 33378539]
72. Mikhelson IV, Bakhtiari S, Elmer TW, Sahakian AV. Remote sensing of heart rate and patterns of respiration on a stationary subject using 94-GHz millimeter-wave interferometry. IEEE Trans Biomed Eng 2011 Jun;58(6):1671-1677. [doi: 10.1109/TBME.2011.2111371] [Medline: 21296700]
73. Adib F, Mao H, Kabelac Z, Katabi D, Miller RC. Smart Homes that Monitor Breathing and Heart Rate. Conference on Human Factors in Computing Systems; 2015 Presented at: 33rd Annual ACM Conference on Human Factors in Computing Systems; April 18-23, 2015; Seoul, Republic of Korea. [doi: 10.1145/2702123.2702200]
74. Guan K, Shao M, Wu S. A remote health monitoring system for the elderly based on smart home gateway. J Healthc Eng 2017;2017:5843504 [FREE Full text] [doi: 10.1155/2017/5843504] [Medline: 29204258]
75. Grijalva CG, Rolfes MA, Zhu Y, McLean HQ, Hanson KE, Belongia EA, et al. Transmission of SARS-COV-2 infections in households - Tennessee and Wisconsin, April-September 2020. MMWR Morb Mortal Wkly Rep 2020 Nov 06;69(44):1631-1634 [FREE Full text] [doi: 10.15585/mmwr.mm6944e1] [Medline: 33151916]

76. Barnett ML, Grabowski DC. Nursing homes are ground zero for COVID-19 pandemic. JAMA Health Forum 2020 Mar 24;1(3):e200369 [FREE Full text] [doi: 10.1001/jamahealthforum.2020.0369]
77. Vanichkachorn G, Newcomb R, Cowl CT, Murad MH, Breeher L, Miller S, et al. Post-COVID-19 syndrome (long haul syndrome): description of a multidisciplinary clinic at Mayo Clinic and characteristics of the initial patient cohort. Mayo Clin Proc 2021 Jul;96(7):1782-1791 [FREE Full text] [doi: 10.1016/j.mayocp.2021.04.024] [Medline: 34218857]
78. Assessment and Testing: Evaluating and Caring for Patients with Post-COVID Conditions: Interim Guidance. Centers for Disease Control and Prevention. 2021 Jun 14. URL: https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/post-covid-assessment-testing.html [accessed 2022-01-15]
79. Natarajan A, Su H, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. NPJ Digit Med 2020 Nov 30;3(1):156 [FREE Full text] [doi: 10.1038/s41746-020-00363-7] [Medline: 33299095]
80. Radin JM, Quer G, Ramos E, Baca-Motes K, Gadaleta M, Topol EJ, et al. Assessment of prolonged physiological and behavioral changes associated with COVID-19 infection. JAMA Netw Open 2021 Jul 01;4(7):e2115959 [FREE Full text] [doi: 10.1001/jamanetworkopen.2021.15959] [Medline: 34232306]
81. op den Akker H, Tabak M, Perianu MM, in 't Veld MHAH, Jones VM, Hofs DHW, et al. 2012 Presented at: 6th International Symposium on eHealth Services and Technologies; July 3-4, 2012; Geneva, Switzerland URL: https://research.utwente.nl/en/publications/development-and-evaluation-of-a-sensor-based-system-for-remote-mo [doi: 10.5220/0004474700190027]
82. Rantz M, Skubic M, Miller S, Krampe J. Using Technology to Enhance Aging in Place. In: Smart Homes and Health Telematics ICOST Lecture Notes in Computer Science. 2008 Presented at: 6th international conference on Smart Homes and Health Telematics; June 28 - July 2, 2008; Ames, IA URL: https://doi.org/10.1007/978-3-540-69916-3_20 [doi: 10.1007/978-3-540-69916-3_20]
83. Mihailidis A, Carmichael B, Boger J. The use of computer vision in an intelligent environment to support aging-in-place, safety, and independence in the home. IEEE Trans Inf Technol Biomed 2004 Sep;8(3):238-247. [doi: 10.1109/titb.2004.834386] [Medline: 15484428]
84. Au-Yeung WM, Miller L, Beattie Z, May R, Cray HV, Kabelac Z, et al. Monitoring behaviors of patients with late-stage dementia using passive environmental sensing approaches: a case series. Am J Geriatr Psychiatry 2022 Jan;30(1):1-11. [doi: 10.1016/j.jagp.2021.04.008] [Medline: 34039534]
85. Jahnke I, Riedel N, Popescu M, Skubic M, Rantz M. Social practices of nurse care coordination using sensor technologies - Challenges with an alert system adoption in assisted living communities for older adults. Int J Nurs Sci 2021 Jul 10;8(3):289-297 [FREE Full text] [doi: 10.1016/j.ijnss.2021.05.011] [Medline: 34307777]
86. Ballering A, Olde Hartman T, Rosmalen J. Long COVID-19, persistent somatic symptoms and social stigmatisation. J Epidemiol Community Health 2021 Feb 23:1 [FREE Full text] [doi: 10.1136/jech-2021-216643] [Medline: 33622802]
87. Raza K. SP0158 Why do patients delay in seeking help. Ann Rheum Dis 2014 Jun 10;73(Suppl 2):42.3-4242. [doi: 10.1136/annrheumdis-2014-eular.6224]
88. Siddeek H, Fisher K, McMakin S, Bass JL, Cortez D. AVNRT captured by Apple Watch Series 4: Can the Apple watch be used as an event monitor? Ann Noninvasive Electrocardiol 2020 Sep 29;25(5):e12742 [FREE Full text] [doi: 10.1111/anec.12742] [Medline: 31994780]
89. Strik M, Caillol T, Ramirez FD, Abu-Alrub S, Marchand H, Welte N, et al. Validating QT-interval measurement using the Apple Watch ECG to enable remote monitoring during the COVID-19 pandemic. Circulation 2020 Jul 28;142(4):416-418. [doi: 10.1161/circulationaha.120.048253]
90. Suresh Kumar S, Dashtipour K, Abbasi QH, Imran MA, Ahmad W. A review on wearable and contactless sensing for COVID-19 with policy challenges. Front. Comms. Net 2021 May 19;2:1. [doi: 10.3389/frcmn.2021.636293]
91. Albahri OS, Albahri AS, Mohammed KI, Zaidan AA, Zaidan BB, Hashim M, et al. Systematic review of real-time remote health monitoring system in triage and priority-based sensor technology: taxonomy, open challenges, motivation and recommendations. J Med Syst 2018 Mar 22;42(5):80. [doi: 10.1007/s10916-018-0943-4] [Medline: 29564649]
92. Kelley B, Fabius R. A path to eliminating $3.6 trillion in wasteful healthcare spending. Thomson Reuters. 2010 Jun. URL: https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.424.548&rep=rep1&type=pdf [accessed 2022-01-15]
93. Mandelin A, Homan P, Shaffer A, Cuda C, Dominguez S, Bacalao E, et al. Transcriptional profiling of synovial macrophages using minimally invasive ultrasound-guided synovial biopsies in rheumatoid arthritis. Arthritis Rheumatol 2018 Jun;70(6):841-854 [FREE Full text] [doi: 10.1002/art.40453] [Medline: 29439295]
94. NIH-Wide Strategic Plan. National Institutes of Health. URL: https://www.nih.gov/about-nih/nih-wide-strategic-plan [accessed 2022-01-15]

Abbreviations

CDC: Centers for Disease Control and Prevention
CPAP: continuous positive airway pressure
EEG: electroencephalogram
FDA: Food and Drug Administration
HIPAA: Health Insurance Portability and Accountability Act

NICE: National Institute for Health and Care Excellence
NIH: National Institutes of Health
PASC: postacute sequelae of SARS CoV-2
POTS: postural orthostatic tachycardia syndrome
SpO2: blood oxygen saturation
WHO: World Health Organization

Edited by C Basch; submitted 06.08.21; peer-reviewed by R Torres, X Cheng, D Banks, T Yang; comments to author 01.10.21; revised version received 15.11.21; accepted 30.11.21; published 26.01.22.

Please cite as:
Corman BHP, Rajupet S, Ye F, Schoenfeld ER
The Role of Unobtrusive Home-Based Continuous Sensing in the Management of Postacute Sequelae of SARS CoV-2
J Med Internet Res 2022;24(1):e32713
URL: https://www.jmir.org/2022/1/e32713
doi:10.2196/32713
PMID:34932496

Original Paper

A Mobile Intervention to Link Young Female Entertainment Workers in Cambodia to Health and Gender-Based Violence Services: Randomized Controlled Trial

Carinne Brody[1], DrPH; Pheak Chhoun[2], MPH; Sovannary Tuot[2,3], MA; Anne E Fehrenbacher[4], PhD; Alexander Moran[5], MScPH; Dallas Swendeman[4], PhD; Siyan Yi[1,2,6], MD, MSc, PhD

[1]Public Health Program, Touro University California, Vallejo, CA, United States

[2]KHANA Center for Population Health Research, Phnom Penh, Cambodia

[3]Department of Community and Global Health, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan

[4]Department of Psychiatry and Biobehavioral Sciences, Semel Institute, University of California Los Angeles, Los Angeles, CA, United States

[5]Department of Epidemiology, University of California Los Angeles, Los Angeles, CA, United States

[6]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore, Singapore

Corresponding Author:
Siyan Yi, MD, MSc, PhD
Saw Swee Hock School of Public Health
National University of Singapore and National University Health System
12 Science Drive 2, #10-01
Singapore, 117549
Singapore
Phone: 65 87533823
Email: siyan@nus.edu.sg

Abstract

Background: Female entertainment workers (FEWs) in Cambodia experience a greater prevalence of human immunodeficiency virus (HIV), other sexually transmitted infections (STIs), psychological distress, substance abuse, and gender-based violence (GBV) than the general female population. Reaching FEWs with health education and linking them to services has been difficult because of their hidden and stigmatized status.

Objective: This study evaluated the efficacy of the Mobile Link intervention in improving FEWs' health by engaging and connecting them to existing HIV, sexual and reproductive health, and GBV services.

Methods: A randomized controlled trial was conducted between March 2018 and June 2019 in the capital city and 3 other provinces in Cambodia. FEWs in the intervention arm received automated twice-weekly Short Message Service messages and voice messages with health information and direct links to outreach workers. The control group received the existing standard care, including free HIV and STI counseling and testing and a toll-free helpline staffed by trained counselors. We used a stratified random sampling method to select participants from 5 study sites in the 4 selected provinces. Initially, we randomly selected 600 participants from a list of 4000 FEWs by age group (18-24 and 25-30 years) and study site using a random number generator and enrolled them in person. The primary outcome measures included self-reported HIV and STI testing, condom use, and contraceptive use assessed through a face-to-face structured interview. We also measured secondary outcomes, including contact with outreach workers, escorted referral service use, forced drinking, and GBV experiences. Intervention effects were modeled using repeated measures, multilevel mixed-effects logistic regression.

Results: A total of 1118 participants were recruited and enrolled in the study. We included 218 FEWs in the intervention arm and 170 FEWs in the control arm in the per protocol analyses after removing 730 dropouts. Evidence of positive intervention effects was detected for the following secondary outcomes: contacting an outreach worker (at 30 weeks: adjusted odds ratio [AOR] 3.29, 95% CI 1.28-8.47), receiving an escorted referral (at 30 weeks: AOR 2.86, 95% CI 1.09-7.52; at 60 weeks: AOR 8.15, 95% CI 1.65-40.25), and never being forced to drink at work (at 60 weeks: AOR 3.95, 95% CI 1.62-9.60). Over time, no significant differences between intervention and control groups were observed for any primary outcomes in the fully adjusted models.

Conclusions: The Mobile Link intervention effectively connected FEWs with outreach workers and escorted referrals but did not show an effect on primary outcomes. Reduced forced drinking at work was also significantly more extensive in the intervention group than in the control group. Longer-term messaging may increase access to services and impact FEWs' health outcomes in the future.

Trial Registration: Clinicaltrials.gov NCT03117842; https://clinicaltrials.gov/ct2/show/NCT03117842

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-018-2614-7

(J Med Internet Res 2022;24(1):e27696) doi:10.2196/27696

KEYWORDS

mHealth; female sex workers; HIV; sexually transmitted infection; linkage to services; sexual and reproductive health; gender-based violence; low- and middle-income countries

Introduction

Female entertainment and sex workers experience a greater prevalence of human immunodeficiency virus (HIV), other sexually transmitted infections (STIs) [1,2], psychological distress [3], substance abuse [3,4], and gender-based violence (GBV) [1,5] than the general female population. These health burdens are intensified by social and structural factors, including poverty, gender inequality, discrimination, and stigmatization [1,2]. In Cambodia, these health trends and associations hold for female entertainment workers (FEWs) employed at establishments such as karaoke bars, restaurants, beer gardens, and massage parlors [6,7]. Approximately one-third of FEWs exchange sex to supplement their income [8]. Transactional sex puts FEWs at high risk for adverse health outcomes, threatening their psychological, physical, and family health [7,8].

FEWs in Cambodia often come from poor and rural families, originally migrating to cities at a young age in hopes of earning higher wages to send to their families [7-9]. FEWs have an HIV prevalence of nearly 10% and low HIV and STI testing rates and service seeking [10]. They also have low rates of modern contraceptive use and high rates of induced abortion [8,11] and are frequent victims of GBV [7,12]. The rates are higher among FEWs younger than 30 years than among the older groups [10-12].

FEWs face barriers to accessing health services. As an HIV key population in Cambodia, they are eligible to receive free health services provided by nongovernment organizations (NGOs). However, many do not seek services due to stigma and discrimination [10]. The population's health risk profile is further exacerbated by Cambodia's 2008 Law on the Suppression of Human Trafficking and Sexual Exploitation, which bans prostitution. Research indicates a strong association between sex work criminalization and sex workers' increased risk for HIV and STIs, sexual and physical violence, and unprotected sex [13]. In Cambodia, outlawing prostitution has amplified the stigma against FEWs and deterred this population from carrying condoms, which the police use as evidence of sex work. The law also creates mistrust of the police among FEWs, which prevents them from reporting instances of GBV to the police, who frequently raid entertainment establishments and arrest FEWs [14,15]. Additionally, the law is associated with a substantial increase in the number of FEWs as more women move from brothels to entertainment venues, where they can continue to exchange sex [6,7]. In 2008, an estimated 13,000 FEWs were reported in Cambodia; that number had risen to approximately 40,000 by 2018 [16].

Health interventions using mobile phones, referred to as mobile health (mHealth), present a viable solution for connecting hard-to-reach, stigmatized, and criminalized populations, such as FEWs, to health services. In recent years, mHealth has received widespread attention due to its applicability in low-resource settings. mHealth has been used effectively in low- and middle-income countries to collect and report community health data [17], disperse health education information [17], raise health awareness [18], and conduct routine check-ins with patients and trigger follow-ups by nurses [19]. However, knowledge gaps persist in mHealth research. Few mHealth interventions have been rigorously evaluated. Much of the existing literature comprises small pilot studies lacking established health indicators and generalizability [20-22]. Additionally, there is a deficiency in mHealth research on interventions targeted toward behavior change and sexual and reproductive health (SRH) [21].

The Mobile Link intervention is an mHealth project aiming to engage and connect young FEWs in Cambodia to existing prevention, care, and treatment services using the automated Short Message Service (SMS) messages and voice messages (VMs). The Mobile Link intervention is based on behavior change theories and years of formative research. The intervention aims to reduce FEWs' risk behaviors and increase HIV, STI, SRH, and GBV service utilization. The project's details are described in the formative study and protocol papers, with an overview provided below [23,24]. This study aims to evaluate the efficacy of the Mobile Link intervention in engaging FEWs; connecting them to existing HIV, SRH, and GBV services; and ultimately improving their health. It strives to fill existing knowledge gaps in mHealth research and provide information about a hard-to-reach, high-risk, and understudied population.

Methods

Trial Design and Settings

The Mobile Link intervention study is a 2-arm multisite 60-week randomized controlled trial (RCT). The trial was conducted in 2 sites in Phnom Penh and 1 site each in Banteay Meanchey, Battambang, and Siem Reap. These provinces were selected

because of substantial populations of FEWs and high HIV burdens.

Participants

The intervention's participant inclusion criteria included (1) being in the age group of 18-30 years; (2) self-identifying as a FEW; (3) working at an entertainment venue in the study sites; (4) being currently sexually active, defined as having engaged in oral, vaginal, or anal sex in the past 3 months; (5) owning a mobile phone; (6) knowing how to retrieve VMs or retrieve and read SMS messages; (7) willing to receive 2 SMS messages/VMs per week for 1 year; (8) providing written informed consent; and (9) agreeing to a follow-up visit after 6 and 12 months.

Randomization

We based our sample size on information from our recent study, which found that 352 of 667 (52.8%) FEWs underwent at least 1 HIV test in the past 6 months [25]. We calculated a sample size that would be able to detect an increase of 13% in HIV testing as a result of the Mobile Link intervention. The sample size was 600 based on a significance level of .05, with 80% power and accounting for 28% attrition. Field workers developed a list of more than 4000 FEWs from the 5 study sites. FEWs in the list were categorized by site and age group (18-24 and 25-30 years old). At each site, 120 FEWs were randomly selected using a random number generator: 60 (50%) in the age group of 18-24 years and 60 (50%) in the age group of 25-30 years. Furthermore, half of the selected FEWs (30 [50%] in the age group of 18-24 years and 30 [50%] in the age group of 25-30 years) were randomly assigned to the intervention arm and the other half to the control arm. Therefore, 60 participants from each arm from each of the 5 sites finally added up to 300 FEWs in the intervention and 300 FEWs in the control arm for a study total of 600 participants.

Recruitment

All participants were recruited in-person at the 5 study sites by trained Mobile Link lay community health workers. Community health workers provided verbal information to FEWs regarding Mobile Link's details because of low literacy rates in this population. Eligible FEWs signed the informed consent form and provided community health workers with mobile numbers for all of their subscriber identification module (SIM) cards and indicated which SIM they used most often. Recruited FEWs were assigned a unique identification number to protect their privacy and blind the researchers from their treatment arm assignment.

At midline, the data collection team recruited additional study participants to replace those who were lost to follow-up by randomly selecting FEWs from the same site and age group on the master list, excluding those who had ever been selected. These replacement participants got an opportunity to participate in the second 30 weeks of the intervention period. We considered participants having at least 2 survey assessments to be active participants in the study.

Patient and Public Involvement

This clinical trial was developed after months of iterative qualitative data collection with FEWs, including 27 focus group discussions (FGDs), 9 in-depth interviews (IDIs), and 2 validation workshops. During this period, participants gave researchers guidance and feedback on the design and delivery of the intervention and recruitment and data collection procedures. Dissemination activities involved a validation workshop, where we presented and discussed findings from the formative studies with representatives of FEWs and key stakeholders, a midterm review workshop, and 2 final findings dissemination workshops.

Intervention

The Mobile Link intervention was informed by behavior change theories and extensive formative research. The intervention provided FEWs with information, resources, and reminders. By utilizing an SMS/VM platform, these services were provided in a convenient, accessible, inexpensive, and confidential manner. Therefore, we theorized that this delivery mechanism would improve FEWs' knowledge of existing resources, risks, risk behaviors, and positive attitudes related to these topics. Increasing knowledge and positive attitudes would contribute to skill acquisition and a positive behavior change.

We conducted a series of formative research activities using participatory methods to create appropriate and relevant health-related messages for FEWs and inform the intervention's development. The formative research process occurred over 6 months. We collected data through FGDs, IDIs, and key informant interviews (KIIs) with venue- and non-venue-based FEWs, in addition to outreach workers and field staff routinely working with this population [23]. Findings from the formative research revealed that FEWs are generally knowledgeable about HIV and STI prevention and transmission. However, they face many structural barriers to optimal health, such as pressure to drink alcohol at work and complicated dynamics of negotiating condom use with clients in a criminalized environment [25,26]. Furthermore, we found that many FEWs face barriers to accessing medical care and services due to stigma, discrimination, and mistreatment from health care workers.

We developed the message-based intervention with the support of local partners InSTEDD iLab and the Women's Media Center (WMC). InSTEDD developed a mobile platform for interactive message delivery and data management using an open source software program. The WMC helped translate messages into Khmer and tailor the contents to be specific, relevant, and engaging, given the cultural context. Example messages included can be found in a previously published paper [23].

After the development, the intervention underwent a 4-week pilot in which 50 FEWs from each study site were randomly selected. The purpose of the pilot was to test whether the platform functioned well with the intervention design and whether the intervention was feasible and acceptable for the participants and other key stakeholders.

The central components of the Mobile Link intervention were the SMS messages and VMs containing health information and referral linkage information to health services and resources.

From the formative research process, 180 messages were designed covering 10 health themes identified as the most important by participants. The health themes covered the following topics: cervical cancer, contraception, general health information, HIV and STI transmission and prevention, miscarriage, pregnancy, alcohol use at work, pregnancy termination, hygiene and vaginal health, and GBV. A message was delivered twice a week for 10 weeks, and the message from each topic area was repeated every 10 weeks for 60 weeks. The health messages were framed using rights-based and health promotion frameworks. Participants could choose to receive the messages via SMS or in VM form that worked with simple and smartphone devices. Those who chose the SMS message option could further personalize their choice by selecting Khmer characters or Romanized Khmer. Each health topic message was followed by a message providing FEWs with the option to be linked to an outreach worker for free. Participants who selected this option were called by Mobile Link's staff, who would provide individualized information over the telephone or face to face and, if needed, would escort the participant to services.

The controls for this study received the existing standard care. Standard care included face-to-face counseling, free HIV and STI testing and condoms, and clinic and hotline phone numbers with a toll-free helpline for clients staffed by trained counselors. The group did not receive the health-related SMS/VM component. However, they received a "check-in" SMS message or VM between baseline and midline and another between midline and endline to stay in touch with the participants and remind them of the interview appointments.

Outcomes and Measures

The primary outcome measures of the Mobile Link intervention were (1) HIV testing, (2) STI testing when experiencing symptoms, (3) contraceptive use, (4) always using condoms with nonpaying partners, and (5) always using condoms with paying partners. The secondary outcome measures were (1) contact with outreach workers, (2) utilization of escorted referrals, (3) forced drinking at work, and (4) responses to GBV and GBV acceptance.

The primary and secondary outcomes were tracked and measured using self-reported data from the baseline, midline, and endline surveys. The survey questionnaires contained items on demographics and background history; entertainment work; sexual behaviors; condom use self-efficacy; HIV risk perception, testing, and treatment; STI testing and treatment; contraception use and pregnancy; GBV and inequity; substance abuse; psychological distress; linkage to health services; and exposure to the Mobile Link intervention. The questionnaires contained approximately 100 questions, which were either dichotomous (eg, yes/no), categorical (eg, type of contraception method used), ordinal (eg, always, frequently, sometimes, and never), or a ratio (eg, number of years working in entertainment venues). The questionnaires were adapted from validated questionnaires used in our previous research on FEWs in Cambodia. The questionnaires were created in English, translated to Khmer, and back-translated to English. The Khmer questionnaires were validated via a pilot test of 15 FEWs with similar characteristics to the intervention participants, who were later excluded from the main surveys.

Data Collection

The data collection process involved a face-to-face baseline, midline, and endline questionnaire survey. The questionnaires were administered in person in Khmer by Mobile Link's female field researchers using the open source Kobo Toolbox software installed on Android-operating tablets. Before data collection, the field researchers underwent a 2-day training in which the questionnaire was also pretested. Field staff connected FEWs to field researchers by making appointments with FEWs, following up the appointments, and guiding FEWs to predetermined interview locations. We conducted the baseline survey before the start of the intervention (March 2018), the midline survey at 6 months after baseline (November 2018), and the endline survey at 12 months after baseline (June 2019). The questionnaires took approximately 25-30 minutes to complete, and the field researchers were blind to the FEWs treatment arm assignment to reduce the possibility of bias.

Ethical Considerations

The Mobile Link intervention engaged community and public health stakeholders to ensure that the study incorporates the best practices and strong ethical standards. Due to the sensitive nature of HIV, SRH, and GBV topics presented in the surveys and questionnaires, additional steps were taken to ensure participants' safety and well-being. First, all data collectors received training related to asking sensitive questions. Second, upon obtaining informed consent, community health workers disclosed information, making clear the sensitive topics discussed in the data collection process. Third, participants were offered escorted referrals to counseling services and provided with services upon request. Participants could be connected to services in the event of an adverse outcome through the SMS/VM platform. In addition, participants could leave the study at any time. Furthermore, participants' identities were kept confidential and stored securely in password-protected files. Coded identifiers were given to participants after obtaining informed consent. No participants' personal identifiers were used in analyses or report writing. Participants received US $5 for compensation of time and transportation reimbursement for their participation. This study was approved by the National Ethics Committee for Health Research (NECHR; no. 142NECHR) within the Ministry of Health in Cambodia and the Touro College Institutional Review Board (no. PH-0117).

Statistical Analyses

STATA/SE 15.1 (College Station, TX, USA) was used for statistical analyses. We tabulated participants' baseline characteristics and distributions of primary and secondary outcome variables for the intervention versus the control arm for the analytic sample (participants with at least baseline and 30-week observation) using frequencies and proportions for categorical variables and means and SDs for continuous variables. These characteristics were compared by group using tests of association, including Pearson chi-square tests of homogeneity for categorical variables and Student *t* tests for continuous variables to ensure the balance between the study

arms. We conducted both crude and cluster-adjusted pooled tests of association to account for clustering within workplace venues. Participant characteristics were then compared for the analytic sample (n=388) versus the non-analytic sample (n=733) to assess significant differences within and between groups for those retained in the study per protocol for at least 2 survey assessments (ie, analytic sample) versus those lost to follow up after the baseline assessment (ie, non-analytic sample).

Intervention effects were assessed using multilevel mixed-effects logistic regression to model all binary outcomes accounting for within-subject correlation from taking repeated measures on the same participants over time (2-level models with observations nested within individuals). Clustered standard errors were computed to account for the similarity of characteristics and behaviors among participants in the same venues and any possible contamination between study arms. Separate models were conducted for each primary and secondary outcome. Model fit was assessed for each outcome using the Akaike information criterion (AIC) and the Bayesian information criterion (BIC).

Predictors in each simple unadjusted 2-level mixed-effects logistic regression model included group, time, and group-by-time interaction terms. Intervention effects for each outcome were determined by group-by-time interaction terms at endline with a significant $P<.05$. Significant interactions indicating intervention effects were graphed using the *marginsplot* command. Midline effects (effects at 30 weeks) are displayed in the figures but not in the tables. For the fully adjusted primary and secondary outcome models, the following covariates were included to control alternative explanations: entertainment job venue type, province, cohabitation, age, and education. For primary outcomes, contact with outreach workers in the past 6 months was also included as a covariate to assess the impact of linkage support on HIV and STI testing, contraceptive use, and condom use.

As a sensitivity analysis, we used intention-to-treat (ITT) principles for modeling primary and secondary outcomes with all participants (n=1121), according to the arm to which they were assigned, and then compared to the results for each outcome from the per protocol modeling with the analytic sample. Per protocol analyses were undertaken to assess the intervention's impact among those who actively participated in the study. Participants lost to follow-up after baseline, resulting in missing outcome data at 6 months, were considered non-users. ITT and per protocol results were consistent for all outcomes regarding the direction, strength, and significance of associations. As such, only the per protocol results are presented in the tables for ease of interpretation.

Protocol Adaptations

There are several protocol deviations to note. The original protocol called for a 12-month (52-week) trial. However, due to high dropout rates at the midterm, we extended the trial to 60 weeks to recruit and enroll more participants who would have the chance to be exposed to the intervention for at least 30 weeks.

We did not anticipate the level of loss to follow-up that occurred and, therefore, did not have a plan in place for replacement recruitment in our original protocol. We decided to recruit replacement participants at midline by randomly selecting FEWs from our master list from the same venue and age group. In our analyses, we defined exposure as having had at least 30 weeks of exposure to the intervention.

Another deviation occurred in our group assignment plan. Initially, we planned to randomize at the entertainment venue level to conduct a cluster RCT. Before the implementation, we changed our trial design to randomize at the individual level due to the high level of movement of FEWs between venues. As a result, we modeled intervention effects using the individual rather than the venue as the analysis unit. We computed clustered standard errors based on the venue rather than including the venue as a level in the mixed-effects outcome models. Finally, we included the venue type (eg, karaoke bar, beer garden, other venues) as a covariate in all our models.

We planned to send out weekly survey questions to intervention participants on various health topics in our study protocol. During intervention development, we heard from pilot participants that they felt reluctant to provide that type of information through the phone. We were also concerned about message fatigue, privacy, and literacy and decided to omit that part of the intervention.

Finally, in our protocol, we stated that we would present an ITT analysis. Because the ITT and per protocol findings were the same, we decided to present the per protocol analysis for ease of interpretation.

Results

The study's participant flow diagram is depicted in Figure 1. Before the intervention started, 3295 FEWs were assessed for eligibility, of whom 828 (25.13%) FEWs did not meet the eligibility criteria, 134 (4.07%) declined to participate, 3 (<0.01%) were lost to follow-up following randomization, and 325 (9.86%) were excluded because of other reasons. Of the included FEWs (2008/3295, 60.94%), 1118 (55.68%) were randomly selected for intervention randomization, of which 435 (38.91%) were allocated to the intervention group and 683 (61.09%) to the control group. By the end of 30 weeks, 217 of 435 (49.9%) FEWs in the intervention and 513 of 683 (75.1%) FEWs in the control group discontinued the study and were replaced. We included 435 (100%) FEWs in the intervention and 683 (100%) FEWs in the control group in intention-to-treat analyses and 218 of 435 (50.1%) FEWs in the intervention and 170 of 683 (24.9%) FEWs in the control group in the per protocol analyses.

Figure 1. Participant flow diagram.

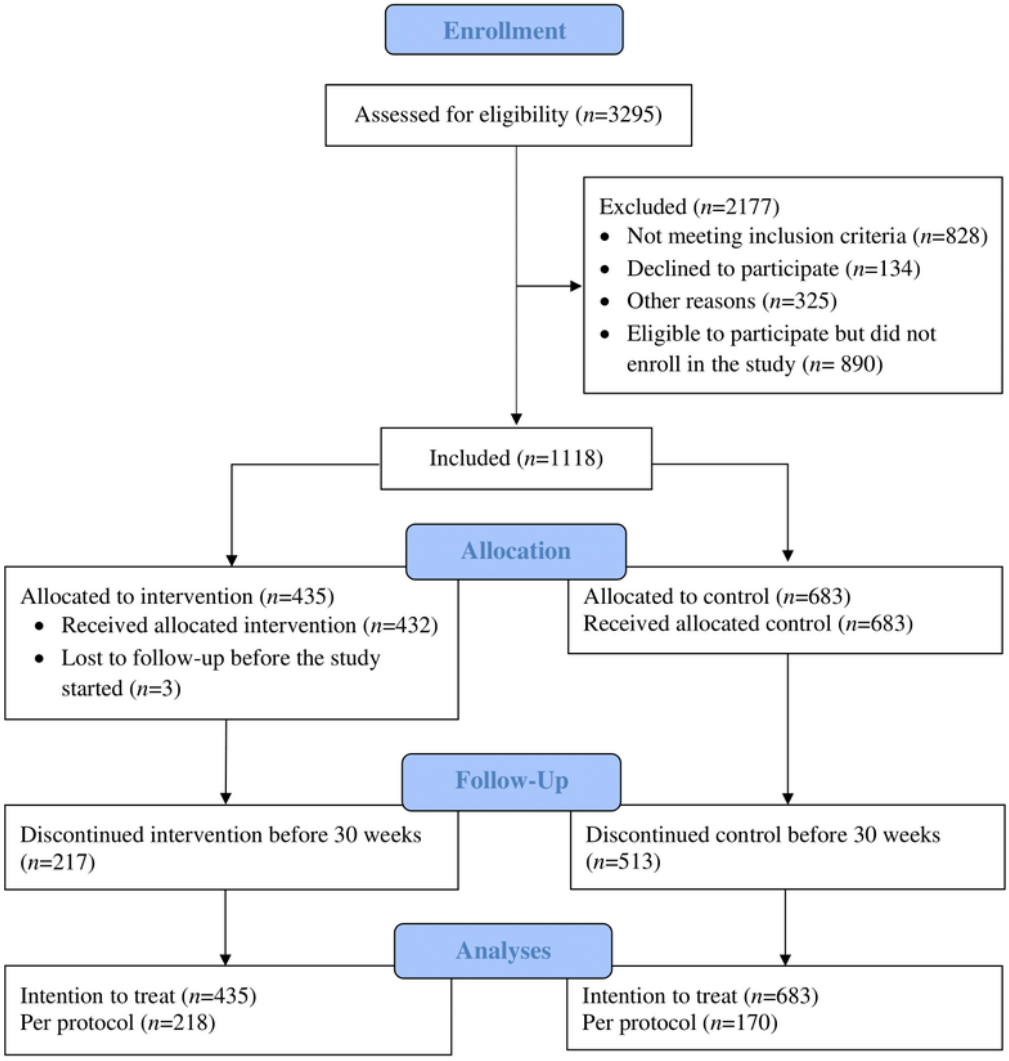


Table 1 displays characteristics of the sample stratified by intervention versus control group with crude and cluster-adjusted pooled association tests. When accounting for clustering by venue, there were no differences in sample characteristics by group at baseline except a significantly lower proportion of currently married participants in the intervention than in the control group (42/218 [19.3%] vs 48/170 [28.2%], *P*=.05). There were crude differences identified by province and entertainment job venue type when not accounting for clustering within venues. The intervention group had a higher proportion of participants from Battambang than the control group (35/218 [16.1%] vs 15/170 [8.8%], *P*=.04). There were marginal but not significant crude differences between the proportions of intervention versus control participants in each of the other 3 provinces. The intervention group had a higher proportion of participants working in karaoke bars (146/218 [66.9%] vs 93/170 [54.7%], *P*=.01). In contrast, the control group had a higher proportion of participants working in beer gardens (26/218 [11.9%] vs 37/170 [21.7%], *P*=.01). All other characteristics were successfully matched between arms, with no statistically significant differences between intervention and control groups.

We compared the characteristics of the analytic (retained) versus non-analytic (lost to follow-up) samples and identified a significant baseline difference in being forced to drink at work. Participants in the analytic sample were more likely to believe that something can be done if a person experiences abuse and to report ever being forced to drink at work. No other baseline differences were identified between analytic and non-analytic samples.

Table 1. Baseline characteristics of female entertainment workers in the analytic sample of the Mobile Link study by intervention and control arms (N=388).

Baseline characteristics	Intervention group (n=218), n (%)	Control group (n=170), n (%)
Demographics		
Age in years, mean (SD)	24.7 (3.8)	24.5 (4.0)
Years of schooling attained, mean (SD)	6.2 (3.0)	6.3 (3.0)
Marital status		
Currently married[a]	42 (19.3)	48 (28.2)
Previously married (widowed, divorced, or separated)	81 (37.2)	55 (32.4)
Never married	95 (43.6)	67 (39.4)
Province		
Phnom Penh	86 (39.4)	51 (30.0)
Battambang[b]	35 (16.1)	15 (8.8)
Banteay Meanchey	52 (23.8)	55 (32.4)
Siem Reap	45 (20.6)	49 (28.8)
Poor as a child (Multidimensional Childhood Poverty Scale score ≥ 3)	190 (87.2)	145 (85.3)
Weekly income in USD, mean (SD)	277.0 (206.2)	272.4 (167.0)
Entertainment job venue type		
Karaoke bar[b]	146 (67.0)	93 (54.7)
Beer garden[b]	26 (11.9)	37 (21.8)
Other (eg, massage parlor, dance club, or freelance in streets/parks)	46 (21.1)	40 (23.5)
Had sex with partner in exchange for money or gifts, past 3 months (intervention n=146 [66.9%], control n=170 [100%])	60 (41.1)	40 (31.5)
Type of message received		
SMS[c] message	99 (45.4)	64 (37.6)
VM[d]	119 (54.6)	106 (62.4)
Primary health outcomes		
Tested for HIV[e] in the past 6 months (intervention n=176 [80.7%], control n=150 [88.2%])	113 (64.2)	95 (63.3)
Tested for STIs[f] when most recently showed symptoms (intervention n=112 [51.4%], control n=71 [41.8%])	27 (24.1)	19 (26.8)
Uses modern contraceptive to prevent pregnancy	68 (31.2)	64 (37.6)
Always uses condom with nonpaying partners	43 (71.1)	31 (77.5)
Always uses condom with paying clients (intervention n=60 [27.5%], control n=40 [23.5%])	23 (22.6)	17 (15.4)
Frequency of forced drinking at work		
Never	123 (56.4)	111 (65.3)
Less than monthly	16 (7.3)	13 (7.6)
Monthly	28 (12.8)	15 (8.8)
Weekly	51 (23.4)	31 (18.2)
Gender-Based Violence Acceptance Scale score (intervention n=26 [11.9%], control n=74 [43.5%]) (range 0-16), mean (SD)	4.4 (3.7)	4.5 (3.2)
Believes you cannot do anything if you or someone you know experiences physical or sexual abuse	60 (27.5)	47 (27.6)
Outreach worker contact only, past 6 months	20 (9.2)	13 (7.7)

Baseline characteristics	Intervention group (n=218), n (%)	Control group (n=170), n (%)
Number of times contacting outreach worker in the past 6 months		
Never	20 (46.5)	21 (58.3)
Once	11 (25.6)	6 (16.7)
2-4 times	11 (25.6)	8 (22.2)
5 or more times	1 (2.3)	1 (2.8)
Received escorted referral		
Received escorted referral for HIV/STI (intervention n=23 [10.6%], control n=23 [13.5%])	7 (30.4)	4 (17.4)
Received escorted referral for vaginal health (intervention n=23 [10.6%], control n=23 [13.5%])	14 (60.9)	16 (69.6)

[a]Indicates a significant difference between intervention and control arms at baseline in a cluster-adjusted test of association by venue (P<.05).

[b]Indicates a significant difference between intervention and control arms at baseline in a crude test of association (P<.05).

[c]SMS: Short Message Service.

[d]VM: voice message.

[e]HIV: human immunodeficiency virus.

[f]STI: sexually transmitted infection.

Table 2 shows intervention effects at endline on primary outcomes in unadjusted and adjusted models (effects at 60 weeks). There was a statistically significant improvement in the frequency of self-reported condom use with nonpaying partners in the control group than in the intervention group in the unadjusted model (effects at 60 weeks: OR 0.26, 95% CI 0.08-0.83). In the intervention group, the proportion of FEWs who reported always using condoms with nonpaying partners was 22.6% (23/102) at baseline, 16.4% (19/116) at midline, and 17.2% (17/99) at endline. Among the control group, 15.5% (17/110) reported this condom use behavior at baseline, 24.8% (30/121) at midline, and 29% (11/38) at endline. However, no significant differences between intervention and control groups over time were observed for any primary outcomes in the fully adjusted models.

Table 2. Intervention effects on primary outcomes in the analytic sample in the unadjusted and adjusted models (N=989). Adjusted models included venue type, province, cohabitation, age, education, and outreach worker contact.

Primary outcomes	OR[a] (95% CI)	AOR[b] (95% CI)
Tested for HIV[c], past 6 months (1=yes, 0=no, n=887 [89.7%])	0.45 (0.18-1.13)	0.40 (0.16-1.04)
Tested for STIs[d], most recent symptoms (1=yes, 0=no, n=394 [39.8%])	1.36 (0.25-7.50)	1.20 (0.21-7.00)
Uses modern contraceptive to prevent pregnancy (1=yes, 0=no, n=989 [100%])	1.06 (0.37-3.02)	0.99 (0.35-2.77)
Always uses condom with nonpaying partners (1=yes, 0=no, n=586 [59.3%])	0.26 (0.08-0.83)[e]	0.50 (0.16-1.58)
Always uses condom with paying clients (1=yes, 0=no, n=242 [24.5%])	1.77 (0.14-22.74)	1.17 (0.08-17.64)

[a]OR: odds ratio.

[b]AOR: adjusted odds ratio.

[c]HIV: human immunodeficiency virus.

[d]STI: sexually transmitted infection.

[e]Indicates a significant difference between intervention and control arms at endline (P<.05).

As shown in Table 3, secondary outcomes with significant intervention effects at endline in unadjusted models included contacting an outreach worker (effects at 60 weeks: OR 3.31, 95% CI 1.06-10.33), receiving an escorted referral (effects at 60 weeks: OR 9.51, 95% CI 2.06-43.95), and never being forced to drink at work (effects at 60 weeks: OR 4.28, 95% CI 1.72-10.65). Statistically significant evidence for positive intervention effects at endline was detected for the following secondary outcomes in the fully adjusted models: receiving an escorted referral (effects at 60 weeks: AOR 8.15, 95% CI 1.65-40.25) and never being forced to drink at work (effects at 60 weeks: AOR 3.95, 95% CI 1.62-9.60). Significant midline effects were also detected for contacting an outreach worker (effects at 30 weeks: AOR 3.29, 95% CI 1.28-8.47) and receiving an escorted referral (effects at 30 weeks: AOR 2.86, 95% CI 1.09-7.52). This effect was not observed at endline for contacting an outreach worker in the adjusted model.

Table 3. Intervention effects on secondary outcomes in the analytic sample in the unadjusted and adjusted models (N=989). Adjusted models included venue type, province, cohabitation, age, and education.

Secondary outcome	OR[a] (95% CI)	AOR[b] (95% CI)
Outreach worker contact, past 6 months (1=yes, 0=no, n=989 [100%])	3.31 (1.06-10.33)[c]	2.82 (0.93-8.55)
Escorted referral, past 6 months (1= yes, 0=no, n=989 [100%])	9.51 (2.06-43.95)[d]	8.15 (1.65-40.25)[d]
Forced drinking at work, past 3 months (1=never, 0=ever, n=989 [100%])	4.28 (1.72-10.65)[d]	3.95 (1.62-9.60)[d]
Believes you can do something if experience abuse (1=yes, 0=no, n=989 [100%])	0.65 (0.27-1.59)	0.68 (0.28-1.62)
GBV[e] (1=high/moderate, 0=low, n=701 [70.9%])	0.81 (0.25-2.59)	0.86 (0.28-2.68)

[a]OR: odds ratio.

[b]AOR: adjusted odds ratio.

[c]Indicates a significant difference between intervention and control arms at endline (P<.05).

[d]Indicates a significant difference between intervention and control arms at endline (P<.01).

[e]GBV: gender-based violence.

Outreach worker contact increased from 19.7% (43/218) at baseline to 31.9% (53/166) at endline in the intervention group and declined from 21.2% (36/170) at baseline to 14.9% (7/47) at endline in the control group. The differences were significant at midline but not at endline. Similarly, escorted referrals increased from 10.6% (23/218) at baseline to 25.9% (43/166) at endline in the intervention group and declined from 13.5% (23/170) at baseline to 6.4% (3/47) at endline in the control group. The differences in escorted referrals were significant at both midline and endline. The proportion of FEWs who reported no forced drinking at work in the past 3 months increased from 56.4% (123/218) at baseline to 70.5% (117/166) at endline in the intervention group and declined from 65.3% (111/170) at baseline to 55.3% (26/47).

Figures 2-4 show significant intervention effects over time in the adjusted models for contact with outreach workers (Figure 2), escorted referrals (Figure 3), and never being forced to drink at work (Figure 4).

Figure 2. Intervention effects on outreach worker contact by timepoint.

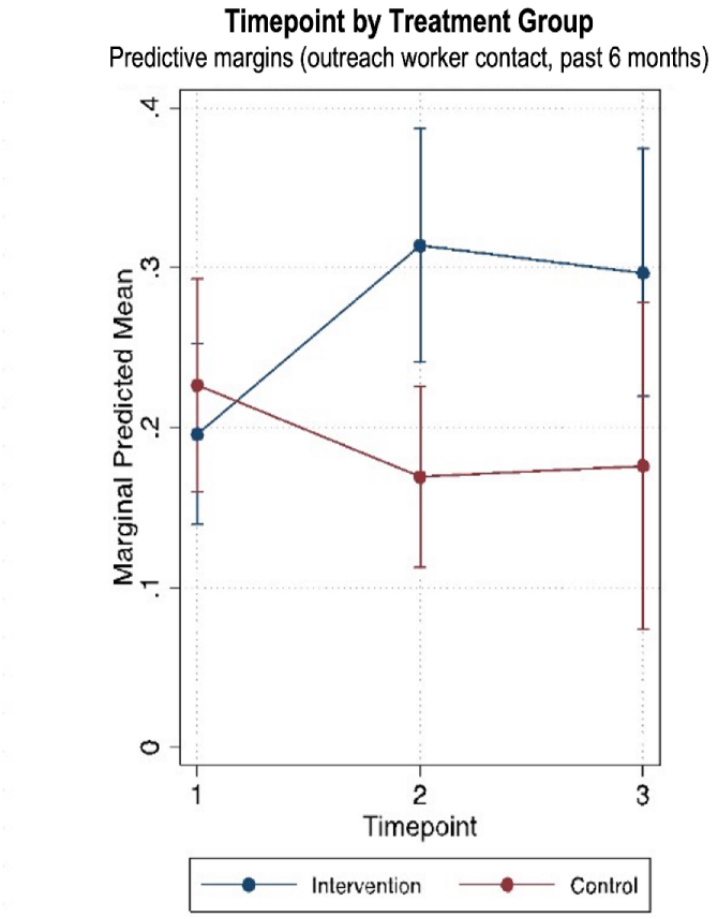


Figure 3. Intervention effects on escorted referrals by timepoint.

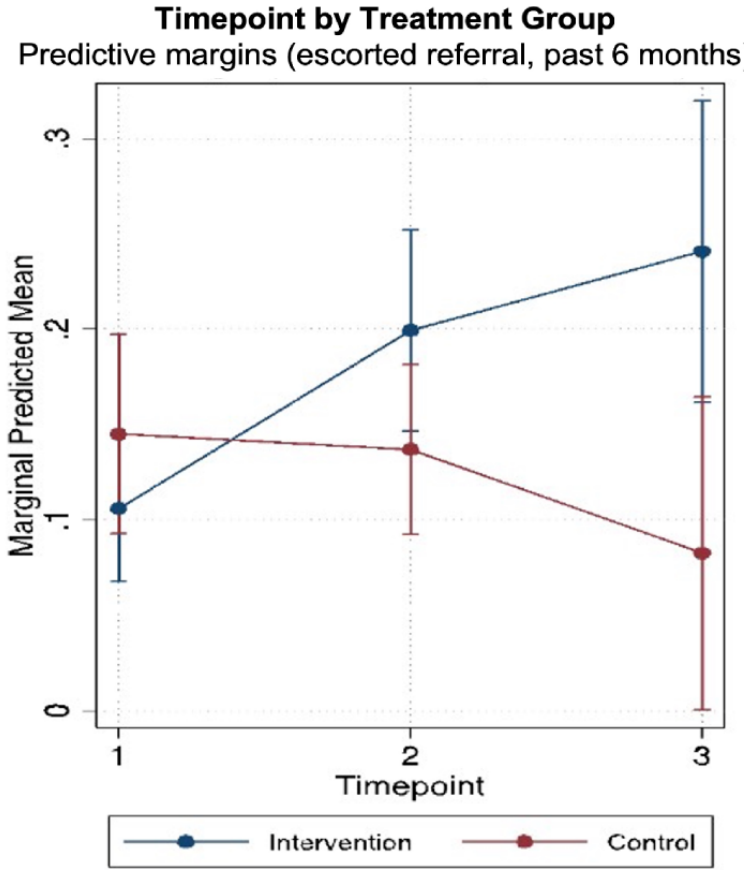
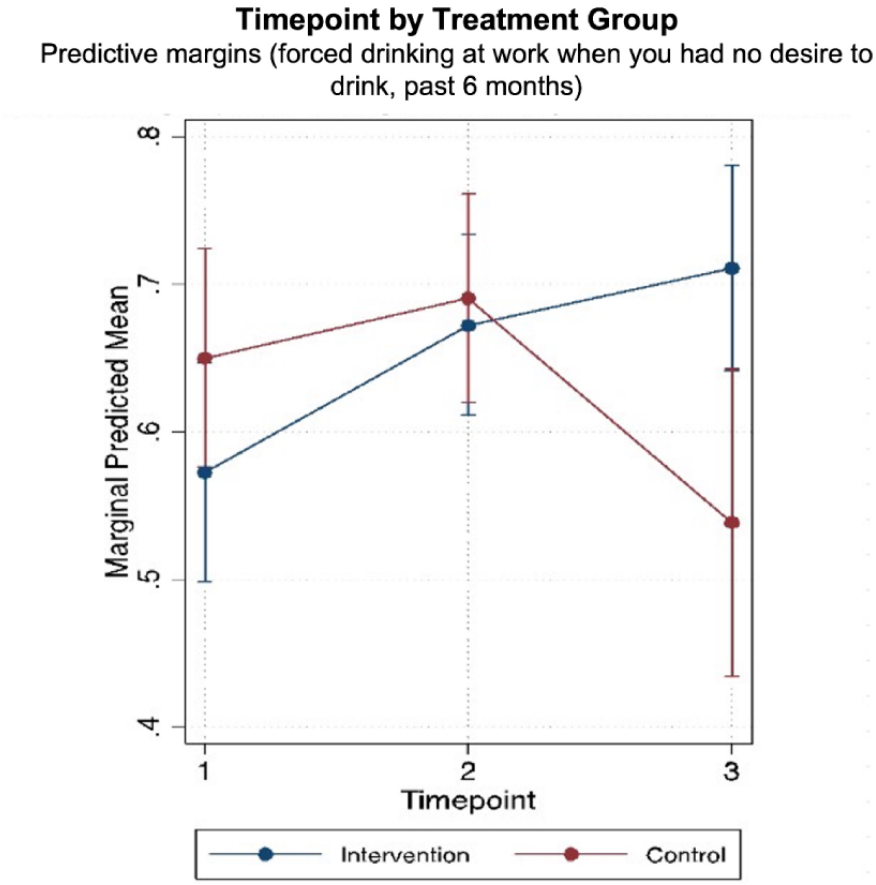


Figure 4. Intervention effects on never being forced to drink at work by timepoint.



Discussion

Principal Results

Our findings suggest that the Mobile Link intervention effectively connects FEWs with outreach workers for health information and escorted referrals. Outreach worker contact increased by 61% in the intervention group and decreased by 30% in the control group. Similarly, escorted referrals increased by 144% in the intervention group and decreased by 53% in the control group over the same period. The reductions in forced drinking at work were larger in the intervention group than in the control group. However, the findings do not indicate an impact on the primary outcome measures, including HIV and STI testing, condom use, and contraceptive use. In addition, improvements on many primary and secondary outcomes in both groups might suggest that enrollment in the study had positive effects on FEWs overall, regardless of whether they were assigned to receive intervention messages.

Our findings indicate that the Mobile Link intervention could reduce the prevalence of forced drinking at work. The proportion of FEWs who reported no forced drinking at work in the past 3 months increased by 25% in the intervention group and decreased by 15% in the control group over the same period. The pressure to drink alcohol at work from supervisors and peers is common for FEWs in Cambodia [25]. During our qualitative assessment for intervention development, FEWs expressed a desire to reduce alcohol consumption at work despite this pressure. Messages developed in response to this request and sent out through the Mobile Link platform advised about subtly avoiding and reducing the effect of heavy drinking. The advices included eating large meals before work, drinking lots of water in between drinks, adding lots of ice to displace the alcohol, and sharing tips to reduce pressure to drink more for tips. These suggestions did not rely on structural-level changes and, therefore, may have been more successfully implemented. As alcohol use is linked to increased sexual risk taking and violence, these findings are promising.

Connecting hard-to-reach populations to prevention and treatment services through outreach workers using mobile phone messages has been shown to be effective in other mHealth studies in low- and middle-income countries [27-29]. RCT evaluations of mHealth interventions have demonstrated positive changes in knowledge and attitudes and, as with our study, have been limited in their ability to show changes in health outcomes [29]. In our theory of change, more health knowledge, paired with more contact with outreach workers, will eventually lead to increased health service use and ultimately improved health outcomes. Our study did not detect changes in health outcomes, perhaps because these changes take longer to occur. It is also possible that several trial implementation challenges may have limited our ability to detect health outcomes changes, including the high loss to follow-up, which was identified as an issue for other mHealth studies in Cambodia [30-32].

Limitations

These findings should be interpreted with regard to study limitations. First, differential loss to follow-up was observed between intervention and control groups, and attrition was exceptionally high among control and replacement participants. This attrition also resulted in a lower overall sample that is underpowered. Second, recruiting new participants to replace those lost to follow-up was conducted according to the initially intended distribution of participants per province rather than where participants were lost, resulting in an overrepresentation of control participants in smaller provinces. Third, high levels of movement between venues among FEWs led to a protocol change, resulting in individual-level rather than venue-level sampling and nonrandom assignment to intervention and control groups. Finally, the participants were not blinded to the intervention. As such, the findings should be interpreted with caution. However, a balance between study arms at baseline was achieved on all primary and secondary outcomes, in both the analytic sample and the full sample, suggesting that intervention and control participants were appropriately matched based on participant characteristics and health behaviors.

Comparison With Prior Work

Despite these limitations, this study demonstrates that mHealth approaches can link hard-to-reach populations to necessary health services. Similar to the implications of meta-analysis findings of SMS interventions for antiretroviral adherence support in sub-Saharan Africa [19], messaging content may not have significant impacts on behaviors relative to linkage to health care workers (ie, nurses or community health workers). However, other text-messaging intervention RCTs focused heavily on message content and with greater intensity but shorter duration than this intervention found significant impacts on multiple HIV-related risk behaviors and relative to live messaging communication with community health workers [31]. Client-centered messaging content may also enhance and sustain engagement in the intervention and seed motivation to connect to a community health worker for support and escorted service linkages. Effects likely depend on the population's risk, resources in context, the intervention intensity, and other intervention details.

Conclusion

Given the positive findings on service linkages for this intervention, we will consider using the Mobile Link model with other key populations in Cambodia and the region. The traditional in-person visitation model by community health workers on 2-week or monthly rotations, the standard of care during this trial, may be enhanced by interventions such as Mobile Link. This study is unique because of the extensive qualitative intervention development period. Replication of any messaging service would benefit from qualitative research to inform adaptation. Successfully linking vulnerable young women to outreach workers and escorted referral services through their mobile phones may lead to linkages to other types of services, events, rights-based information, and behavior change messaging. Longer-term messaging and prompts of community health worker linkage have the potential to increase access to services and may impact FEWs' health outcomes in the future.

Acknowledgments

The research team would like to express gratitude to the participants who contributed their time to participate in the study. We also thank our Cambodian implementing partners, the National AIDS Authority, the National Center for HIV/AIDS, Dermatology and STD, and the Ministry of Post and Telecommunication, for their support. In addition, we would like to acknowledge our funders, the L'Initiative through Expertise France (grant no 16SANIN210).

AEF was supported by a grant from the NIH Fogarty International Center and Office of Behavioral and Social Sciences Research awarded to the University of California Global Health Institute at UCSF (D43TW009343) and a training grant from the National Institute of Mental Health awarded to the UCLA Semel Institute for Neuroscience and Human Behavior (T32MH109205). AEF and DS were supported by a center grant from the National Institute of Mental Health awarded to the UCLA Center for HIV Identification, Prevention, and Treatment Services (P30MH58107).

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 505 KB - jmir_v24i1e27696_app1.pdf]

References

1. Baral S, Beyrer C, Muessig K, Poteat T, Wirtz AL, Decker MR, et al. Burden of HIV among female sex workers in low-income and middle-income countries: a systematic review and meta-analysis. Lancet Infect Dis 2012 Jul;12(7):538-549. [doi: 10.1016/S1473-3099(12)70066-X] [Medline: 22424777]
2. The Joint United Nations Programme on HIV/AIDS (UNAIDS). UNAIDS Guidance Note on HIV and Sex Work. Geneva: UNAIDS; 2009. URL: https://www.unaids.org/sites/default/files/media_asset/JC2306_UNAIDS-guidance-note-HIV-sex-work_en_0.pdf [accessed 2018-06-12]
3. Gorry J, Roen K, Reilly J. Selling your self? The psychological impact of street sex work and factors affecting support seeking. Health Soc Care Community 2010 Sep;18(5):492-499. [doi: 10.1111/j.1365-2524.2010.00925.x] [Medline: 20584087]
4. Scorgie F, Chersich MF, Ntaganira I, Gerbase A, Lule F, Lo Y. Socio-demographic characteristics and behavioral risk factors of female sex workers in sub-saharan Africa: a systematic review. AIDS Behav 2012 May;16(4):920-933. [doi: 10.1007/s10461-011-9985-z] [Medline: 21750918]
5. Argento E, Reza-Paul S, Lorway R, Jain J, Bhagya M, Fathima M, et al. Confronting structural violence in sex work: lessons from a community-led HIV prevention project in Mysore, India. AIDS Care 2011 Jan;23(1):69-74. [doi: 10.1080/09540121.2010.498868] [Medline: 21218278]
6. International Labour Organization (ILO). Cambodia: Addressing HIV Vulnerabilities of Indirect Sex Workers during the Financial Crisis; Situation Analysis, Strategies and Entry Points for HIV/AIDS Workplace Education. Bangkok: ILO; 2011. URL: https://www.ilo.org/wcmsp5/groups/public/---asia/---ro-bangkok/documents/publication/wcms_165487.pdf [accessed 2018-06-12]
7. Brody C, Chhoun P, Tuot S, Swendeman D, Yi S. Childhood conditions, pathways to entertainment work and current practices of female entertainment workers in Cambodia: baseline findings from the Mobile Link trial. PLoS One 2019;14(10):e0216578 [FREE Full text] [doi: 10.1371/journal.pone.0216578] [Medline: 31613881]
8. Ministry of Education, Youth and Sport (MoEYS). Examining Life Experiences and HIV Risks of Young Entertainment Workers in Four Cambodian Cities. Phnom Penh: MoEYS; 2012. URL: https://old.aidsdatahub.org/sites/default/files/publication/Young_entertainment_workers_report_Eng_Final.pdf [accessed 2018-06-12]
9. World Bank Group. Cambodia Economic Update. Phnom Penh: World Bank; 2018. URL: https://documents1.worldbank.org/curated/en/888141543247252447/pdf/132482-WP-PUBLIC-nov-28-Economic-Update-Nov-final-01Low-res.pdf [accessed 2018-06-12]
10. Couture M, Sansothy N, Sapphon V, Phal S, Sichan K, Stein E, et al. Young women engaged in sex work in Phnom Penh, Cambodia, have high incidence of HIV and sexually transmitted infections, and amphetamine-type stimulant use: new challenges to HIV prevention and risk. Sex Transm Dis 2011 Jan;38(1):33-39 [FREE Full text] [doi: 10.1097/OLQ.0b013e3182000e47] [Medline: 21085056]
11. Yi S, Tuot S, Chhoun P, Pal K, Tith K, Brody C. Factors associated with induced abortion among female entertainment workers: a cross-sectional study in Cambodia. BMJ Open 2015 Jul 31;5(7):e007947 [FREE Full text] [doi: 10.1136/bmjopen-2015-007947] [Medline: 26231754]

12. Wieten CW, Chhoun P, Tuot S, Brody C, Yi S. Gender-based violence and factors associated with victimization among female entertainment workers in Cambodia: a cross-sectional study. J Interpers Violence 2020 Aug 17:886260520948145. [doi: 10.1177/0886260520948145] [Medline: 32799728]
13. Platt L, Grenfell P, Meiksin R, Elmes J, Sherman SG, Sanders T, et al. Associations between sex work laws and sex workers' health: a systematic review and meta-analysis of quantitative and qualitative studies. PLoS Med 2018 Dec;15(12):e1002680 [FREE Full text] [doi: 10.1371/journal.pmed.1002680] [Medline: 30532209]
14. United Nations Development Programme (UNDP), Asia-Pacific Regional Centre. Sex Work and the Law in Asia and the Pacific: Laws, HIV and Human Rights in the Context of Sex Work. Bangkok: UNDP, Asia-Pacific Regional Centre; 2012. URL: https://www.unaids.org/sites/default/files/media_asset/JC2306_UNAIDS-guidance-note-HIV-sex-work_en_0.pdf [accessed 2019-01-16]
15. Human Rights Watch. Off the Streets: Arbitrary Detention and Other Abuses against Sex Workers in Cambodia. New York: Human Rights Watch; 2010. URL: https://www.refworld.org/docid/4c45405e2.html [accessed 2018-06-12]
16. Page K, Stein EV, Sansothy NB, Evans J, Couture MC, Sichan K, Young Women's Health Study Collaborative. Sex work and HIV in Cambodia: trajectories of risk and disease in two cohorts of high-risk young women in Phnom Penh, Cambodia. BMJ Open 2013 Sep 10;3(9):e003095 [FREE Full text] [doi: 10.1136/bmjopen-2013-003095] [Medline: 24022389]
17. Agarwal S, Perry HB, Long LA, Labrique AB. Evidence on feasibility and effective use of mHealth strategies by frontline health workers in developing countries: systematic review. Trop Med Int Health 2015 Aug;20(8):1003-1014 [FREE Full text] [doi: 10.1111/tmi.12525] [Medline: 25881735]
18. L'Engle KL, Vahdat HL, Ndakidemi E, Lasway C, Zan T. Evaluating feasibility, reach and potential impact of a text message family planning information service in Tanzania. Contraception 2013 Feb;87(2):251-256. [doi: 10.1016/j.contraception.2012.07.009] [Medline: 22935322]
19. Mbuagbaw L, van der Kop ML, Lester RT, Thirumurthy H, Pop-Eleches C, Ye C, et al. Mobile phone text messages for improving adherence to antiretroviral therapy (ART): an individual patient data meta-analysis of randomised trials. BMJ Open 2013 Dec 17;3(12):e003950 [FREE Full text] [doi: 10.1136/bmjopen-2013-003950] [Medline: 24345901]
20. Chib A, van Velthoven MH, Car J. mHealth adoption in low-resource environments: a review of the use of mobile healthcare in developing countries. J Health Commun 2015;20(1):4-34. [doi: 10.1080/10810730.2013.864735] [Medline: 24673171]
21. Colaci D, Chaudhri S, Vasan A. mHealth Interventions in Low-Income Countries to Address Maternal Health: A Systematic Review. Ann Glob Health 2016;82(5):922-935 [FREE Full text] [doi: 10.1016/j.aogh.2016.09.001] [Medline: 28283147]
22. Gurman TA, Rubin SE, Roess AA. Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. J Health Commun 2012;17 Suppl 1:82-104. [doi: 10.1080/10810730.2011.649160] [Medline: 22548603]
23. Chhoun P, Kaplan KC, Wieten C, Jelveh I, Lienemann M, Tuot S, et al. Using participatory methods to build an mHealth intervention for female entertainment workers in Cambodia: the development of the Mobile Link project. Mhealth 2019;5:24 [FREE Full text] [doi: 10.21037/mhealth.2019.07.02] [Medline: 31559269]
24. Brody C, Tuot S, Chhoun P, Swendeman D, Kaplan KC, Yi S. Mobile Link: a theory-based messaging intervention for improving sexual and reproductive health of female entertainment workers in Cambodia; study protocol of a randomized controlled trial. Trials 2018 Apr 19;19(1):235 [FREE Full text] [doi: 10.1186/s13063-018-2614-7] [Medline: 29673381]
25. Yi S, Tuot S, Chhoun P, Pal K, Chhim K, Ngin C, et al. Sexual behaviors, HIV knowledge, HIV testing attitudes and recent HIV testing among female entertainment workers in Cambodia: a cross-sectional study. PLoS One 2018;13(7):e0198095 [FREE Full text] [doi: 10.1371/journal.pone.0198095] [Medline: 29965968]
26. Brody C, Reno R, Chhoun P, Kaplan K, Tuot S, Yi S. Female entertainment workers and condom use negotiation in post-100% condom use era Cambodia. Arch Sex Behav 2020 Nov;49(8):3065-3074 [FREE Full text] [doi: 10.1007/s10508-020-01649-3] [Medline: 32211995]
27. Ippoliti NB, L'Engle K. Meet us on the phone: mobile phone programs for adolescent sexual and reproductive health in low-to-middle income countries. Reprod Health 2017 Jan 17;14(1):11 [FREE Full text] [doi: 10.1186/s12978-016-0276-z] [Medline: 28095855]
28. Smith C, Ngo TD, Gold J, Edwards P, Vannak U, Sokhey L, et al. Effect of a mobile phone-based intervention on post-abortion contraception: a randomized controlled trial in Cambodia. Bull World Health Organ 2015 Dec 01;93(12):842-50A [FREE Full text] [doi: 10.2471/BLT.15.160267] [Medline: 26668436]
29. Rokicki S, Cohen J, Salomon JA, Fink G. Impact of a text-messaging program on adolescent reproductive health: a cluster-randomized trial in Ghana. Am J Public Health 2017 Feb;107(2):298-305. [doi: 10.2105/AJPH.2016.303562] [Medline: 27997236]
30. Bullen P. Operational challenges in the Cambodian mHealth revolution. JournalMTM 2013;2(2):20-23. [doi: 10.7309/jmtm.2.2.5]
31. Chhim S, Chhea C, Sopheab H, Mathers BM, Kaldor J, Evans JL, et al. Proportion and predictors of loss to follow-up in a longitudinal cohort study of female entertainment and sex workers in Cambodia. Int J STD AIDS 2018 Nov;29(13):1295-1304 [FREE Full text] [doi: 10.1177/0956462418779471] [Medline: 29979143]
32. Reback CJ, Fletcher JB, Swendeman DA, Metzner M. Theory-based text-messaging to reduce methamphetamine use and HIV sexual risk behaviors among men who have sex with men: automated unidirectional delivery outperforms bidirectional

peer interactive delivery. AIDS Behav 2019 Jan;23(1):37-47 [FREE Full text] [doi: 10.1007/s10461-018-2225-z] [Medline: 30006792]

Abbreviations

AIC: Akaike information criterion
AOR: adjusted odds ratio
BIC: Bayesian information criterion
FEW: female entertainment worker
FGD: focus group discussion
GBV: gender-based violence
HIV: human immunodeficiency virus
IDI: in-depth interview
ITT: intention-to-treat
KII: key informant interview
mHealth: mobile health
NGO: nongovernment organization
RCT: randomized controlled trial
SIM: subscriber identification module
SMS: Short Message Service
SRH: sexual and reproductive health
STI: sexually transmitted infection
VM: voice message
WMC: Women's Media Center

Edited by R Kukafka; submitted 03.02.21; peer-reviewed by X Yan, N Khan; comments to author 28.03.21; revised version received 03.05.21; accepted 08.11.21; published 04.01.22.

Please cite as:
Brody C, Chhoun P, Tuot S, Fehrenbacher AE, Moran A, Swendeman D, Yi S
A Mobile Intervention to Link Young Female Entertainment Workers in Cambodia to Health and Gender-Based Violence Services: Randomized Controlled Trial
J Med Internet Res 2022;24(1):e27696
URL: https://www.jmir.org/2022/1/e27696
doi:10.2196/27696
PMID:34982716

Original Paper

Users' Experiences With the NoHoW Web-Based Toolkit With Weight and Activity Tracking in Weight Loss Maintenance: Long-term Randomized Controlled Trial

Elina Mattila[1], PhD; Susanne Hansen[2], MSc; Lise Bundgaard[2], MSc, MA; Lauren Ramsey[3], BS; Alice Dunning[3], PhD; Marlene N Silva[4], PhD; Marja Harjumaa[1], PhD; Miikka Ermes[1], PhD; Marta M Marques[5], PhD; Marcela Matos[6], PhD; Sofus C Larsen[7], PhD; Jorge Encantado[8,9], PhD; Inês Santos[4,10], PhD; Graham Horgan[11], PhD; Ruairi O'Driscoll[3], PhD; Jake Turicchi[3], PhD; Cristiana Duarte[3,6], PhD; António L Palmeira[4], PhD; R James Stubbs[3], PhD; Berit Lilienthal Heitmann[7,12], PhD; Liisa Lähteenmäki[2], PhD

[1]VTT Technical Research Centre of Finland Ltd, Espoo, Finland

[2]MAPP Centre, Aarhus University, Aarhus, Denmark

[3]School of Psychology, Faculty of Medicine and Health, University of Leeds, Leeds, United Kingdom

[4]Centro de Investigação em Desporto, Educação Física, Exercício e Saúde, Faculdade de Educação Física e Desporto, Universidade Lusófona de Humanidades e Tecnologias, Lisbon, Portugal

[5]Comprehensive Health Research Centre, NOVA Medical School, Universidade Nova de Lisboa, Lisbon, Portugal

[6]Center for Research in Neuropsychology and Cognitive Behavioral Intervention, University of Coimbra, Coimbra, Portugal

[7]Research Unit for Dietary Studies, The Parker Institute, Bispebjerg and Frederiksberg Hospital, The Capital Region, Denmark

[8]Centro Interdisciplinar De Estudo Da Performance Humana, Faculdade de Motricidade Humana, Universidade de Lisboa, Lisbon, Portugal

[9]Applied Psychology Research Center Capabilities & Inclusion, ISPA – Instituto Universitário, Lisbon, Portugal

[10]Laboratório de Nutrição, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal

[11]Biomathematics and Statistics Scotland, Aberdeen, United Kingdom

[12]Department of Public Health, Section for General Practice, University of Copenhagen, Copenhagen, Denmark

Corresponding Author:
Elina Mattila, PhD
VTT Technical Research Centre of Finland Ltd
Tekniikantie 21
Espoo, 02150
Finland
Phone: 358 407162230
Email: elina.m.mattila@vtt.fi

Abstract

Background: Digital behavior change interventions (DBCIs) offer a promising channel for providing health promotion services. However, user experience largely determines whether they are used, which is a precondition for effectiveness.

Objective: The primary aim of this study is to evaluate user experiences with the NoHoW Toolkit (TK)—a DBCI that targets weight loss maintenance—over a 12-month period by using a mixed methods approach and to identify the main strengths and weaknesses of the TK and the external factors affecting its adoption. The secondary aim is to objectively describe the measured use of the TK and its association with user experience.

Methods: An 18-month, 2×2 factorial randomized controlled trial was conducted. The trial included 3 intervention arms receiving an 18-week active intervention and a control arm. The user experience of the TK was assessed quantitatively through electronic questionnaires after 1, 3, 6, and 12 months of use. The questionnaires also included open-ended items that were thematically analyzed. Focus group interviews were conducted after 6 months of use and thematically analyzed to gain deeper insight into the user experience. Log files of the TK were used to evaluate the number of visits to the TK, the total duration of time spent in the TK, and information on intervention completion.

Results: The usability level of the TK was rated as satisfactory. User acceptance was rated as modest; this declined during the trial in all the arms, as did the objectively measured use of the TK. The most appreciated features were weekly emails, graphs,

goal setting, and interactive exercises. The following 4 themes were identified in the qualitative data: engagement with features, decline in use, external factors affecting user experience, and suggestions for improvements.

Conclusions: The long-term user experience of the TK highlighted the need to optimize the technical functioning, appearance, and content of the DBCI before and during the trial, similar to how a commercial app would be optimized. In a trial setting, the users should be made aware of how to use the intervention and what its requirements are, especially when there is more intensive intervention content.

Trial Registration: ISRCTN Registry ISRCTN88405328; https://www.isrctn.com/ISRCTN88405328

International Registered Report Identifier (IRRID): RR2-10.1136/bmjopen-2019-029425

(J Med Internet Res 2022;24(1):e29302) doi:10.2196/29302

KEYWORDS

digital behavior change intervention; user experience; technology acceptance; weight-loss maintenance; focus groups; mixed methods; mobile phone

Introduction

Background

Digital behavior change interventions (DBCIs) have the potential to enable scalable solutions for health promotion and disease prevention, and some of them have been found to be effective in weight management [1,2]. The acceptance and use of DBCIs are influenced by several factors. The quality and usability of the DBCI, including the quality and trustworthiness of information and interaction, trust in the privacy and security of the DBCI, and the ease of adoption and use are important for engaging with an intervention [3]. Specific features associated with more active use include frequent updates such as new information being uploaded or new lessons becoming available, and dialogue support features such as reminders and suggestions [4,5]. The inclusion of interaction with a counselor and social interaction have also been found to increase engagement [3-5]. In addition, user- and setting-related factors have an impact on the use of DBCIs, for example, the motivation and agency of the user, the user's personal life situation, and the study setting and recruitment strategies [3,4]. Participants have been found to commit to DBCIs more strongly in randomized controlled trials (RCTs) than in pilot studies and real-life observational studies [4].

The use of a DBCI is required for the users to become exposed to its behavior change mechanisms and gain benefits. The amount and manner of use required to gain benefits may differ between different types of interventions and also between individuals [6,7]. Therefore, it is important to track how users interact with the intervention and their experiences with it to empirically establish and define effective engagement for a new intervention [6].

Technology acceptance models aim to explain and predict user acceptance and adoption of technologies [8]. One of the most widely used models is the Technology Acceptance Model (TAM), originally developed for technologies used in organizational settings [9]. The TAM model consists of 2 main factors that influence whether users will adopt a technology, namely *perceived usefulness*, which is defined as the users' belief that using the system will enhance their job performance, and *perceived ease of use*, which is defined as the users' belief that using the system will be effortless. Different variations and extensions of the TAM model have been developed, including the TAM for Mobile Services (TAMM) [10]. The TAMM adds the dimensions of *perceived ease of adoption* and *trust* and extends the concept of perceived usefulness to *perceived value*. The TAMM model has been previously used for the development and evaluation of DBCIs [11,12].

Traditionally, user acceptance and experiences have been studied in the intervention development phase primarily using qualitative methods. This is an essential part of intervention development to capture the needs and perspectives of the end users [13]. User experiences and the relationship between the user and the technology change over time and in different contexts [14,15]. If a technology is intended for long-term use, the long-term experience determines whether the users will continue to use and recommend the technology to others [15]. Therefore, user experiences should be measured repeatedly during long-term use and in situations where the users are independently using the technologies.

Although quantitative measures of user experience indicate the perceived usability and value, qualitative methods are required to gain a deeper understanding of the reasons associated with user experiences. Focus groups can be used to gather such data from groups of individuals in specific situations. These investigations gather detailed user perspectives on satisfaction or dissatisfaction with services or products that are not obtained by quantitative approaches [16].

Objectives

The NoHoW Toolkit (TK) is a DBCI for WLM (weight loss maintenance) comprising modular intervention content and integrating data from self-assessments, activity trackers, and weight scales. The content of the TK was built on evidence-based theories and techniques of behavior change targeting different psychosocial constructs, which were translated into a digital format [17]. This study employed a mixed methods approach to investigate user experiences during 12 months of TK use. The objectives of this study are as follows: (1) to quantitatively analyze the user experiences of the TK and the changes in user experiences over 1 year; (2) to investigate the use of the TK and its associations with user experience and weight outcomes; and (3) to qualitatively analyze the main strengths, weaknesses, and improvement needs of the TK and the external factors affecting user acceptance of the TK.

Methods

Study Procedures and Materials

Overview

The NoHoW trial was an 18-month, 3-center, 2×2 factorial, single-blind RCT that evaluated the efficacy of the TK in WLM. In total, 1627 participants were recruited from 3 countries—the United Kingdom, Denmark, and Portugal and randomly assigned to 4 arms—(1) control arm (400/1627, 24.58%), (2) *motivation and self-regulation* arm (403/1627, 24.76%), (3) *emotion regulation* arm (416/1627, 25.56%), and (4) *combined* arm (408/1627, 25.07%). The participants were required to be aged ≥18 years, have a verified ≥5% weight loss in the last 12 months with current weight at least 5% below their highest weight, and have had a BMI of ≥25 kg/m^2 before weight loss. Recruitment was conducted through several channels to reach eligible individuals, for example through commercial and municipal weight loss services, registered dieticians and nutritionists, leisure centers, and local and national media coverage and advertisements. The individuals were directed to country-specific recruitment websites and completed a web-based eligibility screener. Eligible individuals were contacted for a telephonic screening interview and provided with study information. Eligible participants were invited to a clinical investigation day where informed consent was obtained before randomization. A detailed description of the study is presented in the paper by Scott et al [18].

Intervention

All the participants received commercial wireless body weight scales (Fitbit Aria [Fitbit LLC]), activity trackers (Fitbit Charge 2 [Fitbit LLC]), and access to the TK website with content tailored to their respective arm. The users also had access to the Fitbit smartphone app (Fitbit LLC) and although they were asked not to use it, access could not be prevented. The TK contained a dashboard and graphs for summarizing the measurements and visualizing long-term progress and enabled simple self-assessments of mood and satisfaction with diet, sleep, activity, and weight as star ratings (1 to 5 stars), and entering free text personal notes into a diary. These parts of the TK were also available to the control arm along with the self-tracking devices. The participants in the 3 intervention arms received intervention content in the form of weekly sessions displayed in the TK as an interactive map. The *motivation and self-regulation* arm had 17 sessions with an estimated minimum time duration of 51.2 minutes (ranging from 20 seconds to 6 minutes and 16 seconds per session), the *emotion regulation* arm had 17 sessions with an estimated minimum time duration of 83.3 minutes (ranging from 40 seconds to 21 minutes and 52 seconds per session), and the *combined* arm had 34 sessions combining the contents of the other 2 intervention arms and having an estimated minimum time duration of 118.3 minutes (ranging from 20 seconds to 12 minutes and 32 seconds per session). Intervention participants were encouraged to complete the intervention sessions during the first 18 weeks of the trial. This was achieved by sending participants weekly emails during this time to introduce the weekly themes and remind them to visit the TK. The control arm also received weekly emails for the first 18 weeks, but they only contained links to generic weight management content. The TK was designed to provide automatic individualized feedback to the *motivation and self-regulation* and *combined* arms based on weight, activity, sleep, and use data. However, owing to an error, the *emotion regulation* arm received the messages instead of the *motivation and self-regulation* arm. The feedback was displayed in the TK as short statements (eg, “your weight management appears better when you are more active”). The TK also provided extra support for weight regain situations (*weight alert*), where an extra module was launched if the user was >3% above their target weight. The TK implementation was fixed for the duration of the trial, that is, no new features or content were added. Only technical errors were rectified when they were reported by the trial staff. A detailed description of the TK is presented by Marques et al [17].

Questionnaires

Quantitative user experience data were collected through electronic questionnaires at 1 (user experience questionnaire at month 1 [UX1]), 3 (user experience questionnaire at month 3 [UX3]), 6 (user experience questionnaire at month 6 [UX6]), and 12 (user experience questionnaire at month 12 [UX12]) months. Table 1 presents the measures in each questionnaire and Multimedia Appendices 1 and 2 contain the detailed questionnaires. The TAMM model was used to create a questionnaire section aimed at measuring the acceptance of the intervention. The TAMM-based items were slightly different in the UX1 from those in the other questionnaires to capture first impression experiences. The System Usability Scale (SUS) was included as a validated measure of the system usability [19]. In addition, items measuring the overall impressions of the TK and its individual features were included. The questionnaires had voluntary open-ended items for providing free-form feedback. Furthermore, the eHealth Literacy Scale (eHEALS) was included in the baseline questionnaire of the RCT to describe the participants’ baseline capacity to engage with eHealth interventions [20].

Table 1. Summary of user experience measures collected through questionnaires.

Measure	Time points, month	Description
eHealth Literacy Scale [20]	0 (Baseline)	eHealth Literacy Scale, consisting of 8 statements measured on a 5-point Likert scale (1=strongly disagree to 5=strongly agree). Higher scores indicate higher digital literacy.
Technology Acceptance Model for Mobile Services, first impressions	1	User acceptance section with 14 items on perceived ease of adoption (3 items), perceived ease of use (4 items), perceived value (4 items), and trust (3 items). Rated on a 5-point Likert scale (1=strongly disagree to 5=strongly agree).
Technology Acceptance Model for Mobile Services, long-term use	3, 6, 12	User acceptance section with 13 items on perceived ease of adoption (2 items), perceived ease of use (3 items), perceived value (5 items), and trust (3 items). Rated on a 5-point Likert scale (1=strongly disagree; 5=strongly agree).
Overall score	1, 3, 6, 12	"What overall score would you give to the service?" Rated on a scale from 0 to 10 (0=very bad; 10=very good).
Recommendation	1, 3, 6, 12	"How likely is it that you would recommend the service to a friend or colleague?" Rated on a scale from 0 to 10 (0=not at all likely; 10=extremely likely).
Intention to continue using the TK[a]	1, 3, 6, 12	"How likely is it that you would consider using the service in the future?" Rated on a scale from 0 to 10 (0=not at all likely; 10=extremely likely).
System Usability Scale [19]	3, 6, 12	System usability measured with 10 items assessed on 5-point Likert scale.
TK feature ratings	3, 6, 12	The 14 main TK features were rated for their importance, ease of use, convenience, enjoyability, satisfaction, and motivation to continue using them on 5-point Likert scales. For the control arm, this section only contained the 5 TK features available to them.
Feedback on the TK	1, 3, 6, 12	Open-ended item: "If you have any other feedback on the TK, you may write it here."
Self-assessment of use	3, 6, 12	"On average, how frequently did you use the TK during the study" and "What option describes your TK use behavior best?" Options are (1) "I used the TK very constantly during the whole study," (2) "I used the TK more in the beginning of the study," (3) "I used the TK more in the end of the study," (4) "I quit using the TK in the middle of the study," and (5) "Other."
Email	3, 6, 12	"I like receiving email from the TK." Rated on a 5-point Likert scale (1=strongly disagree; 5=strongly agree)
Open-ended questions	3, 6, 12	Open-ended items: (1) "What motivated you to continue using the TK," (2) "Why did you use the TK more in the beginning than in the end of the study," and (3) "Why did you quit using the TK in the middle of the study?"

[a]TK: Toolkit.

Focus Group Discussions

Focus group discussions were conducted after 6 months. Focus groups were organized to provide a deeper understanding of the user experiences of the TK. A participant was eligible if they had been using the TK for ≥6 months. In each country, 1 focus group per arm was conducted, leading to a total of 12 groups. Recruitment was done by listing all the participants who had the opportunity to use the TK for ≥6 months at the time of the focus group discussions and recruiting from this list until 8 participants agreed to participate. Participants were recruited via an email that included an invitation letter and an information sheet. Groups were moderated by one or more researchers who were experienced in conducting focus groups and qualitative research and who had not taken part in the TK design or development or participated in delivering the intervention (LB [Denmark], MNS [Portugal], and AD and LR [United Kingdom]). The conversations were guided by a semistructured discussion guide, which highlighted the following 5 main topics of conversation: ways of use, user experience, perceived support in WLM, impact on weight management, and how to improve the TK (Multimedia Appendix 3). Unplanned topics of conversation were also explored based on the issues raised by the participants. Moderators aimed to ensure that no one participant dominated the conversation, and every participant was given the opportunity to contribute to discussions. Ethical approval was granted by the local institutional ethics committees at the Universities of Leeds (17-0082; February 27, 2017), Lisbon (17/2016; February 20, 2017) and the Capital Region of Denmark (H-16030495; March 8, 2017).

Analysis

Quantitative Data Analysis

The TK use was captured in log files, which were used to calculate the total number of visits to the TK and the total duration of use during the 12 months of the study. The percentages of weekly users and retained users were plotted for the control and intervention arms. Retention was determined based on rolling retention, which means that a participant was considered as a retained user if they used the TK during or after a specific week. Intervention completion was calculated for the intervention arm participants based on the duration of time they spent in the sessions assigned to them versus the estimated duration of the sessions. For most sessions, if a user visited a session at least once and remained engaged for at least 33% of the estimated duration, the session was considered as completed. For sessions containing video and audio content, the threshold

was increased, that is, the user was required to spend 50%-80% of the estimated duration in the session. The completion rates were calculated as the percentage of completed sessions compared with the total number of sessions in the arm. The use metrics were reported as medians and IQRs owing to the skewed distributions and nonparametric methods were used for comparisons and correlations (ie, Kruskal-Wallis test, Mann-Whitney *U* test, and Spearman correlation). Spearman correlations between use metrics and weight changes from baseline to 12 months were calculated for each intervention group.

The statistics of responses to the questionnaires were summarized. Quantitative questionnaire sections were descriptively summarized using means and SDs. One-way analysis of variance and Tukey post hoc test were used to determine if there were differences between the intervention arms, and Spearman correlations between the number of visits and the user experience items were calculated. The TAMM-based items were summarized as mean scores over the 4 TAMM dimensions and a total score over all the items. The eHEALS score was calculated as the sum of individual items. The SUS was scored according its guidelines [19]. Principal component analysis was conducted on the 6 items that measured the appeal of the TK features to investigate whether there were underlying or latent variables that accounted for the items. Differences between the centers were investigated by comparing use metrics, eHEALS score, and 3-month user experiences in the control and intervention arms using a general linear model. For this analysis, a logarithmic transformation was applied to the number of visits and total duration of use to normalize their distribution. Quantitative data were analyzed using MATLAB R2017a (Mathworks) and IBM SPSS Statistics version 26 (IBM Corporation). An α level of .05 was used as the threshold for statistical significance.

Qualitative Data Analysis

Focus groups were transcribed and analyzed thematically via an iterative, nonlinear, and nonprescriptive process [21,22]. This involved initial familiarization with the transcripts, reflections on similarities and differences between cases, and systematic coding of the data. After preliminary discussions across the countries on codes and themes, the initial coding framework was developed in Denmark and further iterated collaboratively to integrate the findings from Portugal and the United Kingdom. In each country, experienced researchers were involved in the process and coding was done from original transcripts by native speakers (LB and LL [Denmark], MNS [Portugal], and AD and LR [United Kingdom]). At each stage, decisions on coding and analysis were discussed and revised by all coders and where necessary, the original data sources were revisited to ensure that the decisions were grounded in the data (refer the focus group code book in Multimedia Appendix 4). The analysis was particularly focused on the research questions (ie, the strengths, weaknesses, and improvement needs), while considering unpredicted but relevant themes. Similarly, responses to the open-ended questionnaire items were also thematically analyzed separately but by using the code book of the focus group analysis as the starting point. The analyses were conducted using NVivo (version 12; QSR International).

Results

Response to Questionnaires

On average, the participants responded to the UX1 after 33.7 (SD 33.3) days, to the UX3 after 94.8 (SD 25.6) days, to the UX6 after 196 (SD 22.3) days, and to the UX12 after 378 (SD 31.2) days from the first log in to the TK, which indicated that the timing of the questionnaires was realized according to the plan.

Response rates varied between the questionnaires, with the highest response rate in the UX1 (1096/1627, 67.36%) and the lowest in the UX6 (829/1627, 50.95%). Of the 1627 participants, 1383 (85%) responded to at least 1 of the questionnaires and 427 (26.24%) responded to all the 4 questionnaires, with an approximately even distribution among the arms (110/427, 25.7% respondents in the control arm, 102/427, 23.9% in the *motivation and self-regulation* arm, 114/427, 26.7% in the *emotion regulation* arm, and 101/427, 23.7% in the *combined* arm). The focus groups reached their recruitment target and involved 4.55% (74/1627) participants in total, with a mean of 6.2 (SD 1.3) participants per discussion.

Of the total study sample, 68.65% (1117/1627) were women. The proportion of female participants was higher in the United Kingdom and Denmark (442/555, 79.6% and 429/536, 80%, respectively) than in Portugal (246/536, 45.9%). The mean age of participants was 44.1 (SD 11.9) years, with the oldest participants in Denmark with a mean age of 47.6 (SD 11.5) years, followed by the United Kingdom with a mean age of 44.4 (SD 12.9) years and Portugal with a mean age of 40.1 (SD 9.7) years. The mean BMI was 29.7 (SD 5.3) kg/m^2 at baseline, 30.3 (SD 5.7) kg/m^2 in the United Kingdom, 30.7 (SD 5.3) kg/m^2 in Denmark, and 28.0 (SD 4.5) kg/m^2 in Portugal.

In the UX1, 33.57% (368/1096) of participants responded to the open-ended question. In total, 92.9% (897/966) of participants responded to at least 1 open-ended question in the UX3, 94.2% (781/829) in the UX6, and 97.3% (871/895) in the UX12.

Use Activity

In total, 98.59% (1604/1627) participants logged into the TK at least once. Table 2 summarizes the use metrics in the arms. According to the Kruskal-Wallis test, the total duration of use and completion differed significantly between the arms (P<.001 for both). Post hoc tests showed that the total duration of use in the control arm was significantly lower than that in the other arms (Mann-Whitney *U* test P<.001 for all pairwise comparisons) and that the *combined* arm had a significantly longer duration of use and lower completion percentage than the other intervention arms (Mann-Whitney *U* test P<.001 for both pairwise comparisons).

Figure 1 shows the percentage of users in the control and intervention arms (combined), accessing the TK during each week of the study and the rolling retention for the same period.

The figure shows that although the number of actual weekly users declined rapidly at the beginning, 75.35% (1226/1627) of the users were retained until week 18, that is, the end of the active intervention.

In all the questionnaires, the respondents used the TK significantly more than the nonrespondents (P<.001 for all use metrics in all the questionnaires; data not shown).

The correlations between the TK use and weight outcomes were small, but showed that a higher amount of TK use was associated with weight loss (Table 3). The correlations were strongest in the *combined* arm, followed by the control arm.

Table 2. Use metrics by arm (N=1627).

Arm	Participant, n (%)	Number of visits, median (IQR)	Total duration of use (minutes), median (IQR)	Completion percentage, median (IQR)
Control	400 (24.59)	16 (7-36)	54.5 (21.2-124.1)	N/A[a]
Motivation and self-regulation	403 (24.77)	16 (9-25.75)	146 (77.4-238.4)	76.2 (47.6-90.5)
Emotion regulation	416 (25.57)	15 (9-24)	132.1 (74.3-237.5)	73.7 (47.4-89.5)
Combined	408 (25.08)	17 (10-27.5)	215.4 (112.3-333.2)	64.9 (5.1-83.8)

[a]N/A: not applicable.

Figure 1. Percentage of actual weekly users and retained users according to the rolling retention criterion by study week in the intervention arms and control arm.

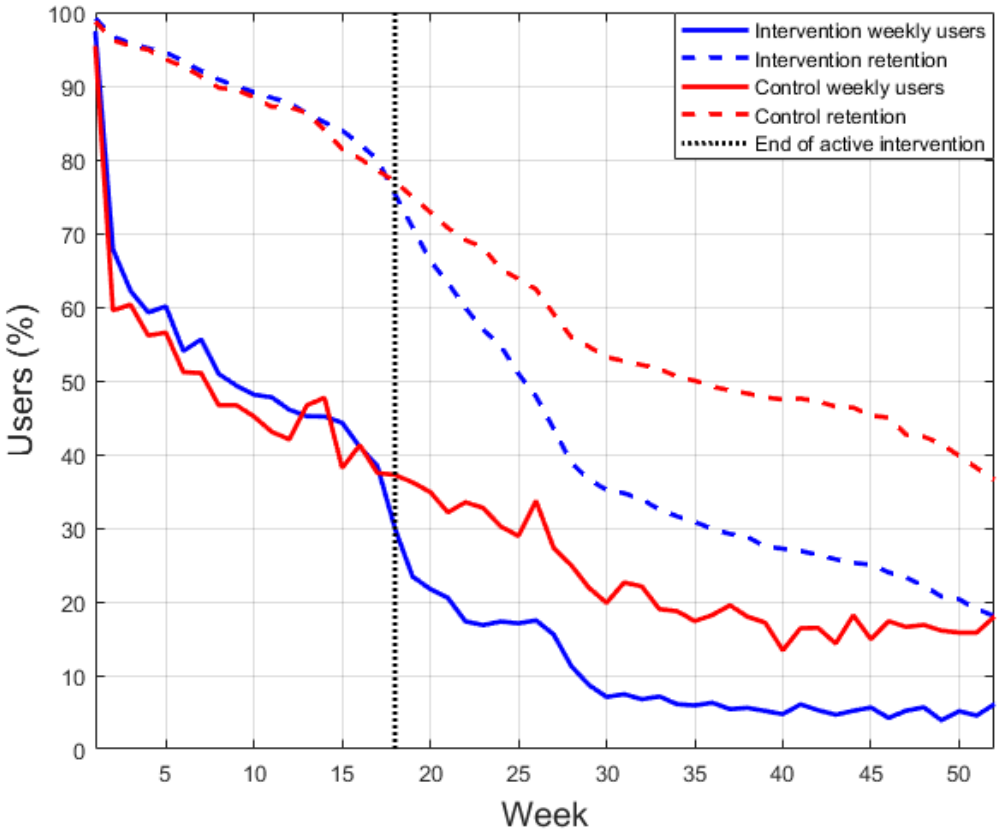


Table 3. Spearman correlations between weight change from baseline to 12 months and use metrics.

Arm	Visits		Total duration		Completion	
	ρ [a]	*P* value	ρ	*P* value	ρ	*P* value
Control	–0.125	.03	–0.114	.04	N/A [b]	N/A
Motivation and self-regulation	–0.083	.14	–0.072	.21	–0.132	.02
Emotion regulation	–0.092	.10	–0.136	.02	–0.106	.06
Combined	–0.198	<.001	–0.175	.002	–0.167	.003

[a]Spearman correlation.

[b]N/A: not applicable.

Quantitative User Experience Results

Digital Literacy

The average eHealth literacy score measured with the eHEALS questionnaire at baseline was 30.8 (SD 5.12; range 10-40) out of a maximum of 40, with higher values signifying higher levels of eHealth literacy. There were no significant differences between the arms and no significant correlations between digital literacy levels and use (data not shown); thus, eHealth literacy is not expected to have affected the results.

Usability

The average SUS scores for measuring the usability of the TK decreased over time in the control arm, from a mean of 70.4 (SD 21.3) in the UX3 to 67.6 (SD 19.9) in the UX12; in the *motivation and self-regulation* arm, from 68.4 (SD 20.9) in the UX3 to 62.1 (SD 20.2) in the UX12; in the *emotion regulation* arm, from 73.0 (SD 19.2] in the UX3 to 67.1 (SD 19.8) in the UX12; and in the *combined* arm, from 69.9 (SD 20.2) in the UX3 to 66.2 (SD 19.5) the UX12. The lowest scores were obtained in the *motivation and self-regulation* arm and they differed significantly from those in the *emotion regulation* arm (*P*=.01) in the UX3 and from all other arms in the UX12 (control arm: *P*=.003; *emotion regulation* arm: *P*=.007; *combined* arm: *P*=.03).

In the control arm, the SUS scores correlated significantly with the number of visits to the TK in all the questionnaires (ρ=0.426 in the UX3, ρ=0.362 in the UX6, and ρ=0.434 in the UX12; all *P*<.001). In the *motivation and self-regulation* arm, there was a small but significant correlation only in the UX3 (ρ=0.142; *P*=.03). In the *emotion regulation* arm, there were significant correlations in all the questionnaires (ρ=0.186, *P*=.003 in the UX3; ρ=0.164, *P*=.02 in the UX6; and ρ=0.218, *P*=.001 in the UX12). In the *combined* arm, there were no significant correlations between SUS scores and number of visits.

Overall Score, Recommendation, and Intention to Continue Using the TK

In most questionnaires, the intervention arm participants gave the TK significantly higher ratings in overall score, likelihood of recommending, and intention to continue to use the TK compared with the control arm participants; ratings declined in all the arms during the study. The ratings mostly did not differ among the 3 intervention arms; only in the UX6, the *combined* arm participants gave a higher overall score than the other intervention arms and had a higher intention to continue than the *emotion regulation* arm (*P*=.01). There were small but significant correlations between the ratings and the number of visits to the TK in the control and *emotion regulation* arms (Table 4).

Table 4. Overall score and ratings for the likelihood of recommending and intention to continue and their correlation with the number of visits.

UX[a] and arm	Overall score		Likelihood of recommending		Intention to continue	
	Score, mean (SD)	ρ [b]	Score, mean (SD)	ρ	Score, mean (SD)	ρ
UX at month 1						
Control	6.74 (2.19)	0.117[c]	6.18 (3.01)	0.125[c]	6.31 (2.82)	0.120[c]
Motivation and self-regulation	7 (2.05)	0.036	7.08 (2.64)	–0.053	6.97 (2.66)	0.034
Emotion regulation	7.3 (1.85)	0.175	7.04 (2.75)	0.134[c]	7.13 (2.53)	0.121[c]
Combined	7.23 (1.86)	0.037	7.15 (2.58)	0.060	7.26 (2.36)	0.079
UX at month 3						
Control	6.21 (2.41)	0.253[d]	5.92 (3.19)	0.143[c]	5.88 (3.04)	0.243[d]
Motivation and self-regulation	6.56 (2.24)	0.114	6.75 (2.87)	0.049	6.67 (2.7)	0.113
Emotion regulation	6.38 (2.34)	0.189[d]	6.46 (2.95)	0.169[d]	6.31 (3.01)	0.183[d]
Combined	6.75 (2.09)	0.003	6.83 (2.78)	–0.015	6.68 (2.8)	0.026
UX at month 6						
Control	5.95 (2.47)	0.262[d]	5.58 (3.27)	0.281[d]	5.39 (3.22)	0.325[d]
Motivation and self-regulation	6.16 (2.22)	0.036	6.22 (3)	0.026	5.84 (2.88)	0.055
Emotion regulation	6.04 (2.44)	0.163[c]	6.01 (3.24)	0.157[c]	5.56 (3.15)	0.210[d]
Combined	6.66 (2.18)	0.103	6.63 (2.89)	0.059	6.32 (2.86)	0.129
UX at month 12						
Control	5.39 (2.46)	0.279[d]	4.62 (3.28)	0.241[d]	4.35 (3.33)	0.263[d]
Motivation and self-regulation	5.9 (2.33)	0.047	5.77 (3.06)	0.010	5.07 (3.12)	0.060
Emotion regulation	5.85 (2.48)	0.169[c]	5.36 (3.37)	0.201[d]	4.8 (3.28)	0.201[d]
Combined	6.1 (2.4)	0.002	5.81 (3.17)	0.020	5.38 (3.12)	0.061

[a]UX: user experience questionnaire.

[b]Spearman correlation between user experience ratings and number of visits to the Toolkit.

[c]Correlation is significant at the significance level of $P<.05$.

[d]Correlation is significant at the significance level of $P<.01$.

TAMM Model

Table 5 presents the mean scores for all the items and the 4 TAMM dimensions. For the total score, the results are presented for all the arms separately, whereas in other results, the intervention arms are combined, as the total score mostly did not differ among the intervention arms. In the UX1, the post hoc tests showed that the *emotion regulation* arm participants gave a higher total score to the TK than the *combined* arm participants ($P=.006$) and the control arm participants ($P=.02$).

The participants of the intervention arms gave a higher rating for *value* in all the questionnaires ($P<.001$ in the UX1, UX6, and UX12 and $P=.01$ in the UX3). The ratings for *ease of adoption*, *ease of use*, and *trust* were mostly similar in the control and intervention arms in all the questionnaires; only in the UX6, the intervention arms had a higher score for *ease of use* ($P=.02$) and in the UX3, for *trust* ($P=.04$). All the ratings decreased during the trial, and only the ratings for *trust* remained above 4 in the intervention arms. In both the control and intervention arms, the highest scores were given to *trust* and the lowest scores to *value*.

When analyzing only the participants who responded to all the user experience questionnaires, the questionnaire averages and the decline during the trial were nearly identical, that is, the decline was not because of the differences in the respondent populations (data not shown).

Table 6 presents correlations between the total number of visits and TAMM total score in all the arms. The control arm showed the highest correlations between the TAMM score and TK use in all the questionnaires.

Table 5. Means and SD of Technology Acceptance Model for Mobile Services (TAMM) scores.

Characteristic and arm	UX1[a]	UX3[b]	UX6[c]	UX12[d]
TAMM total score, mean (SD)				
Control arm	3.83 (0.71)	3.49 (0.76)	3.37 (0.77)	3.22 (0.77)
Motivation and self-regulation	3.80 (0.75)	3.55 (0.73)	3.52 (0.68)	3.36 (0.75)
Emotion regulation	3.97 (0.65)	3.64 (0.70)	3.51 (0.76)	3.37 (0.79)
Combined	3.9 (0.71)	3.6 (0.74)	3.51 (0.74)	3.38 (0.77)
Ease of adoption, mean (SD)				
Control arm	4.06 (0.86)	3.75 (0.95)	3.74 (0.93)	3.64 (1)
Intervention arms	3.98 (0.86)	3.7 (0.85)	3.61 (0.93)	3.52 (0.96)
Ease of use, mean (SD)				
Control arm	4.08 (0.86)	3.58 (0.91)	3.44 (0.92)	3.36 (0.89)
Intervention arms	3.98 (0.87)	3.66 (0.84)	3.6 (0.84)	3.49 (0.88)
Value[e], mean (SD)				
Control arm	3.18 (1)	2.99 (1)	2.77 (1.04)	2.56 (0.98)
Intervention arms	3.5 (0.92)	3.17 (0.93)	3.05 (0.95)	2.86 (0.98)
Trust, mean (SD)				
Control arm	4.13 (0.75)	4.08 (0.77)	4.04 (0.80)	3.91 (0.89)
Intervention arms	4.20 (0.71)	4.19 (0.69)	4.14 (0.71)	4.01 (0.77)

[a]UX1: user experience questionnaires at month 1.

[b]UX3: user experience questionnaires at month 3.

[c]UX6: user experience questionnaires at month 6.

[d]UX12: user experience questionnaires at month 12.

[e]Value is one of the TAMM model dimensions.

Table 6. Spearman correlations between the Technology Acceptance Model for Mobile Services total score and number of visits to the Toolkit.

Arm	UX1[a], ρ	UX3[b], ρ	UX6[c], ρ	UX12[d], ρ
Control	0.193[e]	0.409[e]	0.342[e]	0.379[e]
Motivation and self-regulation	0.142[f]	0.199[e]	0.036	0.059
Emotion regulation	0.126[f]	0.168[e]	0.187[e]	0.247[e]
Combined	0.139[f]	0.065	0.141[f]	0.1

[a]UX1: user experience questionnaires after month 1.

[b]UX3: user experience questionnaires after month 3.

[c]UX6: user experience questionnaires after month 6.

[d]UX12: user experience questionnaires after month 12.

[e]Correlation is significant at the significance level of $P<.01$.

[f]Correlation is significant at the significance level $P<.05$.

Feature Ratings

Principal component analysis was conducted on the 6 appeal items rated on each feature. In all the questionnaires, only 1 factor was found, accounting for 65.2% to 68.9% of the variance in the items in the UX12 and UX3, respectively. The component loadings were similar for all the items, and therefore, the mean of the appeal items was calculated to represent the perception of the features (Table 7).

All the feature ratings were between 3 and 4 on a 5-point scale. At 3 months, the intervention arm participants gave the highest ratings to weekly emails, graphs, and goal setting; at 6 months to graphs, weekly emails, and interactive exercises; and at 12 months, they gave the highest ratings to graphs, interactive exercises, and goal setting. The control arm participants gave the highest ratings to weekly emails and graphs at 3 and 6 months and to graphs and dashboard at 12 months.

Table 7. Ratings for the main features of the Toolkit. Control arm scores are shown only for the features available for them.

Feature and arm	UX3[a], mean (SD)	UX6[b], mean (SD)	UX12[c], mean (SD)
Interactive map	3.52 (1.06)	3.42 (1.03)	3.28 (1.03)
Theme introduction videos	3.67 (1)	3.55 (0.98)	3.41 (1.01)
Text information	3.75 (0.92)	3.66 (0.93)	3.47 (0.91)
Interactive exercises	3.78 (0.96)	3.7 (0.97)	3.55 (0.99)
Audio exercises	3.45 (1.13)	3.4 (1.08)	3.31 (1.1)
Dashboard			
Control arm	3.79 (0.9)	3.64 (0.88)	3.45 (0.95)
Intervention arms	3.78 (0.94)	3.65 (0.95)	3.5 (0.97)
Graphs			
Control arm	3.82 (0.91)	3.72 (0.93)	3.55 (0.96)
Intervention arms	3.88 (0.97)	3.77 (0.96)	3.66 (0.97)
Goal setting	3.85 (0.93)	3.66 (0.97)	3.52 (1.00)
Coping and action plan	3.65 (0.97)	3.51 (0.98)	3.35 (1.03)
Personal notes			
Control arm	3.38 (1.05)	3.17 (1.07)	3.07 (1.12)
Intervention arms	3.28 (1.05)	3.26 (1.02)	3.16 (1.03)
Personal feedback tile	3.42 (1.04)	3.33 (1.02)	3.19 (1.06)
Weight alert	3.78 (1.07)	3.67 (1.06)	3.49 (1.14)
Summary tile			
Control arm	3.57 (1.02)	3.45 (0.91)	3.31 (1.04)
Intervention arms	3.6 (1.03)	3.49 (1.01)	3.36 (1.04)
Weekly emails			
Control arm	3.83 (0.97)	3.68 (1.02)	3.19 (1.16)
Intervention arms	3.94 (0.95)	3.71 (1.02)	3.42 (1.12)

[a]UX3: user experience questionnaires after month 3.

[b]UX6: user experience questionnaires after month 6.

[c]UX12: user experience questionnaires after month 12.

Center Differences

Age and gender were added to the general linear models as covariates, as these background variables differed between the centers. There were no significant differences between the use metrics among the centers (Table 8).

There were differences in the eHEALS scores. The Portuguese intervention participants gave significantly lower scores than the participants in other countries. In addition, the SUS score differed among the centers in the intervention arms, and all other investigated user experience ratings differed among the centers in both the control and intervention arms. In all the user experience ratings, the highest scores were given by Portuguese participants and lowest by the Danish participants.

Table 8. Differences among countries in the control and intervention arms. *P* values are obtained from the general linear model with the adjustment for age and gender.

Characteristic and arm	United Kingdom	Denmark	Portugal	*P* value
Number of visits, median (IQR)				
Control arm	17 (8-34)	20 (8-36)	13 (6.5-44)	.91
Intervention arms	17 (10-26)	17 (9.5-25.5)	14 (7-24)	.12
Total duration in minutes, median (IQR)				
Control arm	54.8 (23.8-140.8)	65.7 (22.8-129.6)	45.7 (13.9-105.5)	.56
Intervention arms	158.7 (81.1-274.3)	169.8 (87.7-280.9)	140.2 (73.6-249.2)	.58
Completion percentage of sessions, median (IQR)				
Control arm	N/A[a]	N/A	N/A	N/A
Intervention arms	71.4 (46-89.5)	76.2 (47.4-89.5)	68.4 (37.8-85.7)	.78
eHEALS[b](range 10-40), mean (SD)				
Control arm	31 (5.1)	31.7 (5.34)	30.1 (4.76)	.27
Intervention arms	31.2 (5.71)	31 (4.8)	29.8 (4.71)	.009
SUS[c] in UX3[d], mean (SD)				
Control arm	70.8 (20.6)	65.8 (20.6)	74.6 (19.6)	.07
Intervention arms	69.3 (20.6)	64.4 (20.5)	77.4 (17.3)	<.001
Overall score in UX3, mean (SD)				
Control arm	6 (2.4)	5.38 (2.57)	7.15 (1.92)	<.001
Intervention arms	6.66 (2.12)	5.72 (2.41)	7.38 (1.78)	<.001
Likelihood of recommending in UX3, mean (SD)				
Control arm	5.39 (3.28)	4.82 (3.1)	7.39 (2.62)	<.001
Intervention arms	6.64 (2.81)	5.69 (3.03)	7.77 (2.3)	<.001
Intention to continue in UX3, mean (SD)				
Control arm	5.59 (3.17)	5.22 (3.12)	6.74 (2.67)	.001
Intervention arms	6.63 (2.8)	5.79 (3.05)	7.30 (2.44)	<.001
TAMM[e] total score in UX3, mean (SD)				
Control arm	3.5 (0.73)	3.29 (0.79)	3.69 (0.71)	.005
Intervention arms	3.63 (0.71)	3.34 (0.7)	3.86 (0.66)	<.001

[a]N/A: not applicable.

[b]eHEALS: eHealth Literacy Scale.

[c]SUS: System Usability Scale.

[d]UX3: user experience questionnaires at month 3.

[e]TAMM: Technology Acceptance Model for Mobile Services.

Experiences From Focus Groups and Open-ended Responses in Questionnaires

Overview

Analysis of the focus groups and open-ended questionnaire data revealed the following four main themes: (1) engagement with features, (2) use decline, (3) external factors affecting user experience, and (4) suggestions for improvements. These reflect the opportunities and challenges regarding the user acceptance of a DBCI. In the following, each theme and its subthemes are described. In the qualitative data, the comments and overall perception of the TK were very similar in the 3 countries and across intervention arms; thus, the results are not divided into arms or countries. Comments from the focus groups are marked with sex, study arm, and country; he survey responses are marked with the code “UX” and the month of the survey is added to the end of the identifier.

Engagement With Features

During the focus groups, participants identified and discussed the features and aspects they found most helpful regarding their WLM and these were also explained in the questionnaire comments.

Support for self-monitoring was perceived positively in both the focus groups and questionnaire comments. For participants, monitoring progress in their weight, sleep, and activity through graphs, dashboards, and weight alerts was identified as very important for their continued WLM, as it gave them a sense of control over their progress. A participant from Portugal commented the following:

> *It is like this, it gave me a conscience, a sense of control of my body that I did not feel, I never felt this before, never....* [Female, emotion regulation arm, Portugal]

Although participants often highlighted the usefulness of the Fitbit app in monitoring weight, they also mentioned that the visualization of their development and progress provided by the TK through graphs and dashboards was important. In contrast, some participants did not find the monitoring as helpful and were reluctant to weigh themselves often, fearing that it would be stressful and potentially detrimental to their WLM:

> *I do not weigh myself every single day, because it gives the wrong impression...You become too stressed.* [Female, control arm, Denmark]

However, in addition to monitoring, the ability of the TK to induce self-awareness and help the participants reflect on their choices was also perceived as important. The idea that everyone must think and know what they do, regarding eating and activity, was fundamental to orient behavior in a different way:

> *And, my moment of reflection of the day was that, [laughter] as I looked at the weight. I thought, I have more or less weight, why? What did I do yesterday? What did I do yesterday that trigger these changes? And I usually do this reflection, while I am dressing for the day I am doing this reflection, humm, how does my body respond to what I ate yesterday?* [Male, combined arm, Portugal]

Participants in the focus groups perceived the TK content and intervention modules as especially useful for their WLM if they prompted them to reflect on the association between behavior and weight. The participants also acknowledged the importance of emotion regulation on WLM, regardless of whether they had it in their intervention, as illustrated by the following comments:

> *...It was important to be able to perceive the association between things...but I felt sad for a while,...why?' 'Why? Do I sleep less than I need? Because am I grumpier? ' 'Because I am doing little exercise and this is also leaving me more tired, or because I am doing a silly food restriction, or because I am overeating and the food is affecting me emotionally; This, I think makes sense….* [Female, motivation and self-regulation arm, Portugal]

> *The biological side of it, about calories and exercise and those things, that's fine. However, for me, there is also the issue of working with your habits, and the impact of emotions.* [Male, motivation and self-regulation arm, Denmark]

Participants further indicated that the TK allowed them to be more conscious of the behavioral patterns linked to sedentary behaviors and bad eating habits and to know themselves better.

Thus, the various features of the TK enabled participants to learn about their own behavior and better understand the antecedents and consequences of their food-related choices, thus helping them self-regulate their behavior and respond to emotional cues.

Use Decline

The use of the TK declined over time. The participants often linked their declining use to various design aspects that they felt hindered their use of the TK. Most often, participants expressed frustration with the login procedure:

> *I think it is cumbersome and illogical that you can't change your username and password.* [Male, control arm, Denmark]

The focus group participants reported varying engagement with the content and sessions of the TK. Although the monitoring tools were important for progress, many of the content-heavy features failed to engage participants in the same way. A prominent issue was the time required to complete the modules as the duration often exceeded the participant's expectations, meaning that the participants had to unwillingly take time out of their day to participate:

> *I think at the beginning it said they [sessions] are only about 5 minutes, you won't have to spend very long on them. And then gradually they got longer until they were about 20 minutes when I was just sitting there listening. I'd have my earphones on but there was still sort of activity going on in the house. It didn't say...even if it came through on the email, it didn't say that in order to do this you need to make sure that you are giving yourself time, space and quietness.* [Female, emotion regulation arm, United Kingdom]

The issue was exacerbated by problems with responsiveness and readability when accessing the TK via a mobile phone, which forced the participants to access it via their home or work computers. This was not always convenient within a busy lifestyle full of interruptions and complications, which was an additional reason for declining use:

> *If it was an app so I could access from my phone, I would use it more often than I do now. At present I have to use it on a laptop or desktop computer which I find quite restrictive in terms of times I can use it.* [UX12, male, emotion regulation arm, United Kingdom]

Although some participants felt that the information provided in the TK was interesting, motivating, and relevant, some felt there was nothing new in the information or that its presentation was not pleasing:

> *They are sometimes a bit patronizing too. I get quite cross at being told to eat my veggies!* [UX6, female, combined arm, United Kingdom]

> *It is boring, and it doesn't tell me anything I didn't already know, so I stopped using it quite early, it is too dry.* [UX3, female, emotion regulation arm, Denmark]

In the questionnaire responses, participants across the intervention arms reported that the weekly modules were tasks that they enjoyed or needed to complete. Weekly reminders were also well received and considered very important in inviting users back to the TK. Without the reminders and the tasks to complete, motivation to use the TK declined:

> *Once the tasks were completed, I kind of forgot about it.* [UX12, female, motivation and self-regulation arm, United Kingdom]

> *It feels a bit like now the weekly support sessions have finished, the momentum for logging into the tool kit has gone...I enjoyed the weekly sessions, found them informative and they helped keep me 'in the zone.* [UX6, female, motivation and self-regulation arm, United Kingdom]

Interestingly, the control arm participants used the TK in a similar way as the intervention arms participants, despite having much less content:

> *First of all it was the prompting from emails from the project and now I use it as a regular part of my weight management. I like to see my progress, good or bad and when it's bad it's a good incentive to get back on track.* [UX6, female, control arm, United Kingdom]

External Factors Affecting the User Experience of the TK

Apart from technical issues, there were several external factors that influenced participants' perceptions of the TK. The activity information collected via the Fitbit smartphone app was available to view on the TK, but the overall design was considered lacking, especially when compared with the Fitbit app, which was also available to all the participants alongside the activity tracker and weight scales. Furthermore, many participants were aware of other commercial weight management apps and used them as a benchmark to measure the TK. This comparison was often unfavorable, and the participants wanted future iterations of the TK to incorporate the best parts of the commercial apps or at least be compatible with them because of their convenience, accessibility, user-friendliness, and design:

> *I have an app on my phone that is more accessible and has a better design, and there is a long way to make the website live up to the same requirements or the same standards. If I did not have the possibility to access the Fitbit app, then I would definitely have looked more at those graphs [on the TK]. Then I would have had to go there to find the same information.* [Male, motivation and self-regulation arm, Denmark]

> *There's nothing particularly wrong with the Toolkit as such - it's just not as nice and easy to use as the Fitbit app. The Fitbit app in conjunction with [other app] provides everything I'd want, all accessible on my phone.* [UX1, female, control arm, United Kingdom]

Finally, study procedures such as the user experience surveys, the food recall questionnaire (INTAKE 24, which was a measurement of the NoHoW trial rather than part of the TK), and visits to the local study sites for weighing and measuring were often confused with the TK and some participants were unsure about what was meant when asked about it:

> *The Toolkit is very good, but the Intake diary is very frustrating.* [UX1, male, emotion regulation arm, United Kingdom]

When asked for feedback on the TK

> *The questionnaires are too long, you get annoyed answering them.* [UX12, male, emotion regulation arm, Denmark]

Improvement Needs

The focus group participants had many ideas to improve the TK. The most prevalent suggestion was the need to create an app, not only to make the service more competitive but also to circumvent issues with login and access and make it easier for the participants to prioritize and plan when to use it. The possibility of personalizing the TK according to one's own use and interests was also highlighted. Similarly, the participants wanted the content to be more tailored to their specific needs, for example, information about the expected time allocation required for the activities and designing a tool that provides immediate help in tempting food situations:

> *Making it into an app would make it more aggressive, which I think in our busy lives we could all do with reminding...If you could have a more aggressive app...it would remind you and say, this is what you have asked for, this is what you want. Sometimes, personally I need reminding.* [Female, emotion regulation arm, United Kingdom]

> *...as a Toolkit user, eventually, I could have the opportunity to change the initial [Map] panel, to access all information but only having my 'favorites' on the initial panel...To rearrange it the way I want it.* [Male, control arm, Portugal]

The lack of interaction with other participants was also regarded as a weakness of the TK. Creating a social platform to allow the participants to interact with others, discuss their progress, and offer support when needed was considered important. The participants spoke about previous experiences in which social support, knowledge exchange, shared activities, and friendly competition had a positive impact on their weight loss:

> *I'm thinking more from a personal motivation point of view and support, while we are all undertaking the programme, wouldn't it be great for us to interact and support each other.* [Female, emotion regulation arm, United Kingdom]

Furthermore, many participants asked for more feedback and notifications from the system. Although the TK provided individualized feedback in 2 arms, it was not widely discussed

in the focus groups. One reason for this might be insufficient data to provide personalized feedback owing to the participants' low level of engagement. The prompting emails sent from the TK seemed to be the main trigger for use and once the reminder messages stopped, use declined. However, the users wished they had received periodic personalized feedback emails with advice on how to reach the individual goals and believed that they would have helped them sustain their engagement:

> *...one monthly [email], one was enough, one monthly. Because, lately, I have not been contacted, since I came here the last [time] I did not receive any more emails, I do not know; I've already wondered, sometimes, what was I going there for, does the project still exist?* [Male, motivation and self-regulation arm, Portugal]

> *...depending on our use, of the results you are having, [hum-hum] humm, some feedback on that, i.e. you are getting these results, maybe a few suggestions of what you can do to continue towards the goals that we are defining for ourselves. This could be an important feedback.* [Male, combined arm, Portugal]

In addition, some participants expressed confusion as the content on the website was not updated during the trial period. Their use of other commercial services created the expectation of regular updates:

> *We have never seen an update; a year has passed and not one update was made...* [Male, motivation and self-regulation arm, Portugal]

In addition, there was some confusion among the participants regarding what data would be accessible to researchers. Although this issue was not explicitly discussed, the participants often thought that the researchers had access to all their data, which may have impacted the participants' engagement with and use of the TK, as not receiving feedback from researchers was seen as unfair:

> *At the moment it seems like they must be collecting a huge amount of data, which is for the research program, but actually the people that are forming that data are not getting any of that.* [Female, motivation and self-regulation arm, United Kingdom]

Discussion

Principal Findings

We investigated the quantitative and qualitative user experiences and objectively measured the use of the TK, a digital WLM intervention, during the NoHoW RCT. Our goal was to investigate the specific factors affecting user experiences and use of the TK over a 1-year period.

TK use was high during the 18-week active intervention, when >75.35% (1226/1627) of the participants were retained, and >31.71% (516/1627) accessed the TK on a weekly basis. These rates are comparable with the high retention rates found for smartphone-based health interventions, ranging from 29% to 79.6% [23]. Later, when the users were free to use the TK at their own discretion, a rapid decline in use was seen, especially in the intervention arms. A similar effect was reported by Mattila et al [24]. For the TK, the most likely reason for the decline was the discontinuation of the email reminders and many participants felt that they had completed their tasks in the intervention. The users completed between 64.9% (*combined* arm) and 76.2% (*motivation and self-regulation* arm) of their assigned intervention sessions.

A higher amount of use was associated with 12-month weight loss in all the intervention arms with small but significant correlations. The metrics with the highest correlation differed among the arms. In the control arm, where the TK only contained self-monitoring–related features, the number of visits had the highest correlation with weight loss. In the *motivation and self-regulation* arm, which consisted of modules with textual content and interactive exercises, intervention completion had the strongest correlation with weight loss. In the *emotion regulation* arm, which consisted of significant video and audio content, the total duration of use had the highest correlation with weight loss. Finally, the *combined* arm, which combined the content of the other 2 arms, showed significant correlations for all the use metrics, with the strongest correlation being between the number of visits and weight loss. This highlights the need to investigate the TK use from different perspectives and by using different metrics. Donkin et al [7] previously reported that in physical health interventions, the number of logins was most consistently associated with effectiveness, probably because of the high emphasis on self-monitoring, whereas intervention completion was most related to effects of psychological health interventions. In this study, both types of metrics were found to be associated, which is relevant because the intervention consisted of both psychological content and self-monitoring. However, it must be noted that to draw conclusions on the effectiveness of the TK for weight loss, a more detailed analysis needs to be conducted by considering the potential confounders such as self-weighing frequency.

The eHealth literacy scores were compared with those reported in previous studies [25]. Overall, the score did not correlate with use and was not expected to affect TK use. The SUS scores measuring the usability of the TK ranged from 73 to 62.1 and were at their highest after 3 months of use. The scores were similar to the average SUS scores found in a large collection of studies and would correspond to a grade C, that is, satisfactory or average level of usability [26]. The *motivation and self-regulation* arm yielded the lowest usability scores. This may be because their intervention content comprised several interactive exercises compared with the *emotion regulation* arm that mainly comprised video and audio content and the control arm that comprised only the dashboard and graphs. Generally, user experience ratings decreased in all the arms during the 12-month follow-up period. The intervention arm participants gave significantly higher ratings to the TK regarding overall score, likelihood of recommending, intention to continue, and value (a TAMM dimension) compared with the control arm; otherwise, the control and intervention arm experiences were very similar. There were also very few differences among the intervention arms. In the TAMM dimensions, the perceived ease of use and perceived value ratings of the intervention arms were at a similar level as in the self-directed intervention in the

study by Ma et al [27] and the decline during long-term use was of a similar magnitude. A more intensive coach-led intervention did not show similar declines in these dimensions [27].

The correlations between user experiences and use were relatively small, but they were consistently significant in the control and *emotion regulation* arms. There were some differences among the countries. The Portuguese participants in the intervention arms had significantly lower eHealth literacy scores than participants in other countries. However, interestingly, they were also the most positive in their user experiences, both in the control and intervention arms.

According to the qualitative findings, a major reason for deteriorated use and user experiences, were the technical difficulties, such as difficulties in logging in and the TK's suboptimal performance on mobile devices that hindered its use. Although content and delivery could not be changed during the RCT, these findings highlight the need to reserve more time for iterative development and testing with target users and on different devices before starting a trial. The users were also disappointed by the lack of new and updated content in the TK and the absence of reminders after 18 weeks. It was decided early in the planning stages that there would be no updates to the TK or the content to avoid confounding of the results during the follow-up period. Furthermore, all the participants were given a Fitbit tracker and weight scales as part of the trial and although they were asked not to use the associated app, many of them used it and compared it with the TK, often unfavorably. These findings illustrate that, in the current digital health environment, even research-stage DBCIs need to be able to compete with commercial apps in their technical functioning and design. Resourcing of DBCI development and trial design need to allow for continuous content and design updates to avoid solutions becoming outdated before the end of the trial.

External factors also influenced the user experience and use of the TK. The users spoke about lifestyle-related barriers, such as being busy at work, going on a holiday, or having illnesses. These barriers have also been reported previously for mobile health apps [28]. DBCIs operate in challenging and uncontrolled environments and they should be able to anticipate and adapt to life events and have mechanisms to cope with them; for example, enable users to mute them for some time and then gently pull the user back to the service, especially in case of suspected relapse, as also suggested by Mattila et al [29].

Of the features, the reminder emails and self-monitoring–related tools (especially graphs, goal setting, and plans) were the most favorably experienced. Self-monitoring is one of the key behavior change techniques in weight management and is pervasive in both weight management treatments and commercial weight management apps [30,31]. Users are accustomed to and expect this feature; therefore, well-designed self-monitoring should be incorporated in DBCIs, while acknowledging that it may not work for everyone and that it may act through different mechanisms that are not yet fully understood [32]. The responsive features, especially the weight alerts, were also among the best-liked features.

Although generally enjoyed, there were mixed feelings regarding the intervention content, especially the emotion regulation content, which is a more novel approach. Although many participants appreciated it, some did not understand it, perhaps because it did not meet their expectations on what a WLM intervention should comprise. In contrast, some participants in the *motivation and self-regulation* arm expressed missing the emotion regulation aspect. These findings suggest that the approach and content of interventions should be personalized according to individual needs and preferences to ensure better acceptance.

Easier access through a smartphone app was most often mentioned as a technical improvement need. A social platform was also desired, as many participants had previous experiences with peer support helping in their weight management and hoped the TK would facilitate contact and engagement with peers. This feature was purposefully left out of the TK as it could have introduced uncontrolled or conflicting components to the intervention and biased the results.

Finally, the TK did not always match user expectations, which caused dissatisfaction and confusion. A clear introduction to the intervention approaches, requirements regarding devices and time allocation, and expected ways of use should be provided. Furthermore, as many participants received the trial procedures mixed with the TK, it is difficult to assess whether some of the negative experiences were related to the burden of trial procedures or the TK itself.

Strengths and Limitations

The strengths of this study include a mixed methods design, large sample size, and focus groups conducted in different countries. In total, 85% (1383/1627) of the total study sample provided user experience responses at some point during the trial. However, we cannot ignore that the results represent the views of the more active users of the TK. Thus, it is not possible to predict the way in which this has skewed the findings. It is likely that active users used the TK more often because they had a positive experience. However, it is also possible that they were committed to using the TK regardless of their experience and were more exposed to the negative sides of the TK. Indeed, an association between negative user experience and more frequent use has been reported earlier by Tuch et al [14].

Similarly, focus groups are likely to include participants at the high end of activity and involvement with the intervention, which may also bias the findings. However, steps were taken to limit this bias. Participants were recruited using a ranking system, rather than using a direct convenience sample of participants who volunteered to participate.

Future Work

Data-driven methods are expected to enable deeper personalization, adaptivity, and reactivity in DBCIs. For example, they enable the prediction, prevention, or detection of lapses in weight management behaviors [33,34]. Data from this trial may enable identification of specific profiles regarding weight maintenance and identification of the type of intervention content that is most effective or acceptable to the participants.

It is important to design DBCIs considering their intended way of use and make this clear to the user, that is, how often and for

how long the intervention should be used. Here, the users were not always sure how they were supposed to use the TK after the active intervention. This study highlighted that continuous interaction with users is important and reminders are crucial in sustaining use. Reminder and interaction schemes should be designed to ensure effective engagement, but bearing in mind that making users dependent on an intervention is usually not a desirable goal [6].

Furthermore, the measurement of long-term user experience should be included as an integral part of DBCIs to guide further development, updates, and adaptation of the intervention. Similar to other aspects of an intervention, user experience varies between individuals and one solution is probably not optimal or pleasing for all.

Conclusions

In the NoHoW RCT, most TK users were retained during the active intervention spanning the first 18 weeks of the study, when they were emailed weekly and reminded to access the intervention content. Technical difficulties related to access and use of the TK and the lack of new and updated content in TK hindered its use and user experiences. Personalized features, including the reminder emails, the self-monitoring–related tools, and weight alerts were well received and rated. These results highlight that the target users of DBCIs are already accustomed to using different types of existing health apps and services and are quick to compare new services with them. This emphasizes the need to finalize and evaluate the user experience and appearance of new services extensively before launching large-scale trials and to allow and enable updates of the content and design during the trial to avoid becoming outdated. Failure to provide an intervention with adequate application development, acceptance, and usability may undermine the original aims of the trial (in this case, testing the impact of self and emotion regulation strategies). Thus, DBCIs should be built to be more responsive and adaptive by monitoring the engagement, progress, and signs of relapse and reacting to those promptly to achieve continued sufficient level of engagement.

Future trials should focus not only on participants' experience of mobile health devices but also on the full user experience of the interventions to ensure that the best possible support and experience is offered.

Acknowledgments

The authors thank Sarah E Scott for her valuable contributions as the trial manager and in the user experience evaluation, and Susana Cunha for her contribution in conducting and reporting the focus groups. This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 643309. The material presented and views expressed here are the responsibility of the authors only. The European Union Commission does not take responsibility for any use made of the information set out.

Authors' Contributions

RJS, BLH, ME, and GH conceived the NoHoW project. CD, ALP, and SCL were the local trial managers. MH, EM, ME, and RJS designed the quantitative user experience evaluation. LL compiled the focus group protocol. EM, SH, and LL drafted the manuscript. EM conducted the use and quantitative user experience analyses. SH, LB, LR, AD, and MNS facilitated focus groups in different countries, coded the transcripts, and collaborated in the qualitative analysis and write-up of the focus group results. All the authors reviewed and approved the final manuscript.

Conflicts of Interest

RJS consults for Slimming World through Consulting Leeds, which is a wholly owned subsidiary of the University of Leeds. Slimming World was a former partner in NoHoW. MMM and GH have previously consulted for Slimming World. The other authors have no conflicts of interest to declare.

Multimedia Appendix 1
User experience questionnaire at month 1: first impressions questionnaire.
[PDF File (Adobe PDF File), 74 KB - jmir_v24i1e29302_app1.pdf]

Multimedia Appendix 2
User experience questionnaires at months 3, 6, and 12. The user experience questionnaire was repeated at 3, 6, and 12 months.
[PDF File (Adobe PDF File), 184 KB - jmir_v24i1e29302_app2.pdf]

Multimedia Appendix 3
Structured discussion guide.
[PDF File (Adobe PDF File), 109 KB - jmir_v24i1e29302_app3.pdf]

Multimedia Appendix 4
Focus group code book.

[PDF File (Adobe PDF File), 145 KB - jmir_v24i1e29302_app4.pdf]

Multimedia Appendix 5
Consort eHealth Checklist.
[PDF File (Adobe PDF File), 11210 KB - jmir_v24i1e29302_app5.pdf]

References

1. Holmes WS, Moorhead SA, Coates VE, Bond RR, Zheng H. Impact of digital technologies for communicating messages on weight loss maintenance: a systematic literature review. Eur J Public Health 2019 Apr 01;29(2):320-328. [doi: 10.1093/eurpub/cky171] [Medline: 30239699]
2. Kozak AT, Buscemi J, Hawkins MA, Wang ML, Breland JY, Ross KM, et al. Technology-based interventions for weight management: current randomized controlled trial evidence and future directions. J Behav Med 2017 Feb;40(1):99-111. [doi: 10.1007/s10865-016-9805-z] [Medline: 27783259]
3. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
4. Kelders SM, Kok RN, Ossebaard HC, Van Gemert-Pijnen JE. Persuasive system design does matter: a systematic review of adherence to web-based interventions. J Med Internet Res 2012;14(6):e152 [FREE Full text] [doi: 10.2196/jmir.2104] [Medline: 23151820]
5. Brouwer W, Kroeze W, Crutzen R, de NJ, de VNK, Brug J, et al. Which intervention characteristics are related to more exposure to internet-delivered healthy lifestyle promotion interventions? A systematic review. J Med Internet Res 2011;13(1):e2 [FREE Full text] [doi: 10.2196/jmir.1639] [Medline: 21212045]
6. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
7. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
8. Taherdoost H. A review of technology acceptance and adoption models and theories. Procedia Manuf 2018;22:960-967. [doi: 10.1016/j.promfg.2018.03.137]
9. Davis FD. A technology acceptance model for empirically testing new end-user information systems : theory and results. Thesis (Ph. D.)--Massachusetts Institute of Technology, Sloan School of Management. 1986. URL: https://dspace.mit.edu/handle/1721.1/15192 [accessed 2021-12-03]
10. Kaasinen E. User acceptance of location-aware mobile guides based on seven field studies. Behav Inform Technol 2005 Jan;24(1):37-49. [doi: 10.1080/01449290512331319049]
11. Kaasinen E, Mattila E, Lammi H, Kivinen T, Välkkynen P. Technology acceptance model for mobile services as a design framework. In: Human-Computer Interaction and Innovation in Handheld, Mobile and Wearable Technologies. IGI Global. Hershey, Pennsylvania: IGI Global; 2011:80-107.
12. Heikkilä P, Mattila E, Ainasoja M. Field study of a web service for stimulating the positive side of stress: entrepreneurs' experiences and design implications. BMC Med Inform Decis Mak 2019 Oct 28;19(1):200 [FREE Full text] [doi: 10.1186/s12911-019-0909-6] [Medline: 31660962]
13. Yardley L, Morrison L, Bradbury K, Muller I. The person-based approach to intervention development: application to digital health-related behavior change interventions. J Med Internet Res 2015;17(1):e30 [FREE Full text] [doi: 10.2196/jmir.4055] [Medline: 25639757]
14. Tuch A, Trusell R, Hornbæk K. Analyzing users' narratives to understand experience with interactive products. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. USA: ACM; 2013 Presented at: CHI '13: CHI Conference on Human Factors in Computing Systems; April 27 - May 2, 2013; Paris France p. 2079-2088. [doi: 10.1145/2470654.2481285]
15. Kujala S, Roto V, Väänänen-Vainio-Mattila K, Karapanos E, Sinnelä A. UX Curve: a method for evaluating long-term user experience. Interact Comput 2011 Sep;23(5):473-483. [doi: 10.1016/j.intcom.2011.06.005]
16. Stewart DW, Shamdasani PN. Focus Groups: Theory and Practice. Thousand Oaks, CA: SAGE Publications; Sep 2014:1-224.
17. Marques M, Matos M, Mattila E, Encantado J, Duarte C, Teixeira P, et al. Development of a theory and evidence-based digital intervention tool for weight loss maintenance: the NoHoW Toolkit. J Med Internet Res 2021 Dec;23(12):e25305. [doi: 10.2196/preprints.25305]
18. Scott SE, Duarte C, Encantado J, Evans EH, Harjumaa M, Heitmann BL, et al. The NoHoW protocol: a multicentre 2×2 factorial randomised controlled trial investigating an evidence-based digital toolkit for weight loss maintenance in European adults. BMJ Open 2019 Sep 30;9(9):e029425 [FREE Full text] [doi: 10.1136/bmjopen-2019-029425] [Medline: 31575569]

19. Brooke J. SUS: A 'Quick and Dirty' usability scale. In: Jordan PW, Thomas B, Weerdmeester BA, McClelland AL, editors. Usability Evaluation in Industry. London: Taylor & Francis; 1996.
20. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
21. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006 Jan;3(2):77-101. [doi: 10.1191/1478088706qp063oa]
22. Attride-Stirling J. Thematic networks: an analytic tool for qualitative research. Qual Res 2016 Nov 07;1(3):385-405. [doi: 10.1177/146879410100100307]
23. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
24. Mattila E, Orsama A, Ahtinen A, Hopsu L, Leino T, Korhonen I. Personal health technologies in employee health promotion: usage activity, usefulness, and health-related outcomes in a 1-year randomized controlled trial. JMIR Mhealth Uhealth 2013;1(2):e16 [FREE Full text] [doi: 10.2196/mhealth.2557] [Medline: 25098385]
25. Richtering SS, Hyun K, Neubeck L, Coorey G, Chalmers J, Usherwood T, et al. eHealth literacy: predictors in a population with moderate-to-high cardiovascular risk. JMIR Hum Factors 2017 Jan 27;4(1):e4 [FREE Full text] [doi: 10.2196/humanfactors.6217] [Medline: 28130203]
26. Lewis JR. The System Usability Scale: past, present, and future. Int J Hum–Comput Interact 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
27. Ma J, Xiao L, Blonstein AC. Measurement of self-monitoring web technology acceptance and use in an e-health weight-loss trial. Telemed J E Health 2013 Oct;19(10):739-745 [FREE Full text] [doi: 10.1089/tmj.2013.0009] [Medline: 23952787]
28. Ahtinen A, Mattila E, Väätänen A, Hynninen L, Salminen3 J, Koskinen E, et al. User experiences of mobile wellness applications in health promotion: user study of Wellness Diary, Mobile Coach and SelfRelax. In: Proceedings of the 3d International ICST Conference on Pervasive Computing Technologies for Healthcare. 2009 Presented at: 3d International ICST Conference on Pervasive Computing Technologies for Healthcare; April 1-3, 2009; London, UK. [doi: 10.4108/icst.pervasivehealth2009.6007]
29. Mattila E, Korhonen I, Salminen JH, Ahtinen A, Koskinen E, Särelä A, et al. Empowering citizens for well-being and chronic disease management with wellness diary. IEEE Trans Inf Technol Biomed 2010 Mar;14(2):456-463. [doi: 10.1109/TITB.2009.2037751] [Medline: 20007055]
30. Semlitsch T, Stigler FL, Jeitler K, Horvath K, Siebenhofer A. Management of overweight and obesity in primary care-A systematic overview of international evidence-based guidelines. Obes Rev 2019 Sep;20(9):1218-1230 [FREE Full text] [doi: 10.1111/obr.12889] [Medline: 31286668]
31. Bardus M, van Beurden SB, Smith JR, Abraham C. A review and content analysis of engagement, functionality, aesthetics, information quality, and change techniques in the most popular commercial apps for weight management. Int J Behav Nutr Phys Act 2016;13(1):35 [FREE Full text] [doi: 10.1186/s12966-016-0359-9] [Medline: 26964880]
32. Hartmann-Boyce J, Boylan A, Jebb SA, Aveyard P. Experiences of self-monitoring in self-directed weight loss and weight loss maintenance: systematic review of qualitative studies. Qual Health Res 2018 Jul 01:124-134. [doi: 10.1177/1049732318784815] [Medline: 29984630]
33. Helander EE, Vuorinen A, Wansink B, Korhonen IK. Are breaks in daily self-weighing associated with weight gain? PLoS One 2014;9(11):e113164 [FREE Full text] [doi: 10.1371/journal.pone.0113164] [Medline: 25397613]
34. Forman EM, Goldstein SP, Zhang F, Evans BC, Manasse SM, Butryn ML, et al. OnTrack: development and feasibility of a smartphone app designed to predict and prevent dietary lapses. Transl Behav Med 2019 Mar 01;9(2):236-245 [FREE Full text] [doi: 10.1093/tbm/iby016] [Medline: 29617911]

Abbreviations

DBCI: digital behavior change intervention
eHEALS: eHealth Literacy Scale
RCT: randomized controlled trial
SUS: System Usability Scale
TAM: Technology Acceptance Model
TAMM: Technology Acceptance Model for Mobile Services
TK: Toolkit
UX1: user experience questionnaire at month 1
UX3: user experience questionnaire at month 3
UX6: user experience questionnaire at month 6
UX12: user experience questionnaire at month 12
WLM: weight loss maintenance

Edited by R Kukafka; submitted 04.06.21; peer-reviewed by CI Sartorão Filho, A Kozak; comments to author 17.08.21; revised version received 19.09.21; accepted 14.10.21; published 10.01.22.

Please cite as:
Mattila E, Hansen S, Bundgaard L, Ramsey L, Dunning A, Silva MN, Harjumaa M, Ermes M, Marques MM, Matos M, Larsen SC, Encantado J, Santos I, Horgan G, O'Driscoll R, Turicchi J, Duarte C, Palmeira AL, Stubbs RJ, Heitmann BL, Lähteenmäki L
Users' Experiences With the NoHoW Web-Based Toolkit With Weight and Activity Tracking in Weight Loss Maintenance: Long-term Randomized Controlled Trial
J Med Internet Res 2022;24(1):e29302
URL: https://www.jmir.org/2022/1/e29302
doi:10.2196/29302
PMID:35006081

Original Paper

Web-Based Problem-solving Training With and Without Peer Support in Veterans With Unmet Mental Health Needs: Pilot Study of Feasibility, User Acceptability, and Participant Engagement

Kyle Possemato[1], PhD; Justina Wu[2], MPH; Carolyn Greene[3], PhD; Rex MacQueen[1]; Daniel Blonigen[4], PhD; Michael Wade[1], PhD; Jason Owen[2], PhD; Terence Keane[5], PhD; Deborah Brief[5], PhD; Steven Lindley[4], PhD; Annabel Prins[2], PhD; Margaret-Anne Mackintosh[2], PhD; Eve Carlson[2], PhD

[1]Veterans Affairs Center for Integrated Healthcare, Syracuse, NY, United States

[2]National Center for Post Traumatic Stress Disorder, Veterans Affairs Palo Alto Health Care System, Palo Alto, CA, United States

[3]Veterans Affairs Office of Mental Health Services and Suicide Prevention, Washington, DC, United States

[4]Veterans Affairs Palo Alto Healthcare System, Palo Alto, CA, United States

[5]National Center for Post Traumatic Stress Disorder, Veterans Affairs Boston Healthcare System, Boston, MA, United States

Corresponding Author:
Eve Carlson, PhD
National Center for Post Traumatic Stress Disorder
Veterans Affairs Palo Alto Health Care System
3801 Miranda Avenue
Palo Alto, CA, 94304
United States
Phone: 1 6507143064
Email: eve.carlson@va.gov

Abstract

Background: eHealth tools have the potential to meet the mental health needs of individuals who experience barriers to accessing in-person treatment. However, most users have less than optimal engagement with eHealth tools. Coaching from peer specialists may increase their engagement with eHealth.

Objective: This pilot study aims to test the feasibility and acceptability of a novel, completely automated web-based system to recruit, screen, enroll, assess, randomize, and then deliver an intervention to a national sample of military veterans with unmet mental health needs; investigate whether phone-based peer support increases the use of web-based problem-solving training compared with self-directed use; and generate hypotheses about potential mechanisms of action for problem-solving and peer support for future full-scale research.

Methods: Veterans (N=81) with unmet mental health needs were recruited via social media advertising and enrolled and randomized to the self-directed use of a web-based problem-solving training called Moving Forward (28/81, 35%), peer-supported Moving Forward (27/81, 33%), or waitlist control (26/81, 32%). The objective use of Moving Forward was measured with the number of log-ins. Participants completed pre- and poststudy measures of mental health symptoms and problem-solving confidence. Satisfaction was also assessed post treatment.

Results: Automated recruitment, enrollment, and initial assessment methods were feasible and resulted in a diverse sample of veterans with unmet mental health needs from 38 states. Automated follow-up methods resulted in 46% (37/81) of participants completing follow-up assessments. Peer support was delivered with high fidelity and was associated with favorable participant satisfaction. Participants randomized to receive peer support had significantly more Moving Forward log-ins than those of self-directed Moving Forward participants, and those who received peer support had a greater decrease in depression. Problem-solving confidence was associated with greater Moving Forward use and improvements in mental health symptoms among participants both with and without peer support.

Conclusions: Enrolling and assessing individuals in eHealth studies without human contact is feasible; however, different methods or designs are necessary to achieve acceptable participant engagement and follow-up rates. Peer support shows potential for increasing engagement in web-based interventions and reducing symptoms. Future research should investigate when and for

whom peer support for eHealth is helpful. Problem-solving confidence should be further investigated as a mechanism of action for web-based problem-solving training.

Trial Registration: ClinicalTrials.gov NCT03555435; http://clinicaltrials.gov/ct2/show/NCT03555435

(J Med Internet Res 2022;24(1):e29559) doi:10.2196/29559

KEYWORDS

problem-solving training; mHealth; peer specialists; veterans

Introduction

Background

Many Americans have mental health needs that go unmet. A national survey found that 11.8 million US adults report unmet mental health needs, with about half of these respondents reporting that they received some, but not enough, mental health services to meet their needs, and about half reporting that they received no mental health services [1]. Even in health care systems that screen for mental health problems and offer treatment, such as the Veterans Health Administration (VHA) and the Department of Defense (DoD), there are substantial numbers of individuals with unmet mental health needs [2,3]. In a study of veterans who screened positive for depression and posttraumatic stress disorder (PTSD) and were offered mental health care, less than half received any mental health treatment [4]. Barriers to treatment engagement include concerns with stigma, a desire to be self-reliant, a lack of appeal of face-to-face therapy, and the inconvenience of traveling to an appointment during the day [5,6]. More recently, health and safety concerns surrounding the COVID-19 pandemic emerged as new barriers to traditional psychotherapy [7].

Novel approaches are needed to engage individuals with unmet needs in mental health care. Electronic or *eHealth* tools are increasingly being used, given their ability to reach many users at relatively low costs, and growing evidence shows that many individuals prefer to use eHealth as part of their health care [8,9]. eHealth could dramatically diminish access problems related to travel distance and time, wait times for appointments, financial burden, stigma, and desire for self-reliance [10].

Websites and mobile apps that help individuals manage their mental health concerns are now widely available, and multiple meta-analyses have described their benefits for users [11,12]. One such program is Moving Forward (MF), a free educational and life coaching program that is a web-based adaptation of problem-solving therapy. MF was built by VHA in partnership with the DoD as part of a coordinated public health initiative to help veterans and service members who have difficulties with mental health. It was first made publicly available in 2011, and a program evaluation found that 750 veterans accessed it in 2020. Problem-solving therapy is an evidence-based, transdiagnostic cognitive behavioral treatment that has been successfully adapted to multiple contexts, including problem-solving training for primary care patients [13]. The intervention has a robust evidence base for a variety of disorders and is among the recommended treatments for depression and suicide prevention in the VHA/DoD Clinical Practice Guidelines. Problem-solving therapy is particularly well-suited to individuals who see their distress as a result of life problems rather than a mental health problem and want to solve their problems themselves.

The benefits gained from eHealth interventions tend to vary by how much an intervention is used, how well users apply the learned strategies to their daily lives, and users' confidence in their ability to manage their mental health problems, known as mental health self-efficacy [14,15]. High rates of attrition in eHealth are ubiquitous, leaving many users to experience no benefits from the intervention [16]. On the basis of these findings, the supportive accountability model was developed to guide how human coaching can increase adherence to eHealth interventions. This model predicts higher engagement in eHealth when the user feels accountable to a coach who is seen as trustworthy and benevolent and having relevant expertise [17].

The unique characteristics of peer specialists may allow them to be particularly effective as eHealth coaches. Peer specialists are individuals who have lived experience with medical or mental health conditions, have benefitted from treatment, can model healthy behaviors, and can provide emotional support. A recent systematic review of 30 studies found that peer support (PS) of eHealth interventions has strong potential for clinical effectiveness; however, more research was recommended to investigate whether and how PS affects user engagement in eHealth and what mechanisms are related to better health among users [18]. The VHA has made a large investment in its peer workforce. Veteran peers now assist other veterans in coping with medical and mental health conditions, including providing support for the VHA's suite of mental health apps and websites [19]. Research is needed to understand the best ways to deliver PS of eHealth tools and how processes to connect veterans with eHealth and peers can be automated.

Objectives

This study is designed to test novel methods of delivering care and generate hypotheses for future research on eHealth delivery. The first aim of this study is to test the feasibility and user-perceived acceptability of the research and intervention methods. Specifically, we test the feasibility of an automated electronic system for recruitment, enrollment, intervention delivery, and data collection in a national sample of veterans recruited via social media. We test user-perceived acceptability via participant engagement in MF and peer sessions, as well as user satisfaction. The second aim is to examine engagement in the web-based MF course for those with and without PS. We hypothesize that the use of MF would be greater for the PS condition than for the self-directed (SD) condition. The third aim is to conduct an exploratory examination of changes in problem-solving skills and problem-solving confidence as

possible mechanisms of action to increase engagement in MF or reduce mental health symptoms. The fourth aim is to conduct an exploratory examination of the association of PS sessions with changes in mental health symptoms. This study adds to the early research on how PS can enhance engagement in eHealth and makes unique contributions by investigating a fully automated system for research and interventions and exploring potential mechanisms of action.

Methods

This study was found to meet all human subjects, data security, and privacy requirements for research approval by the local institutional research boards where the study investigators were located. The ClinicalTrials.gov identifier is NCT03555435.

Participants and Procedures

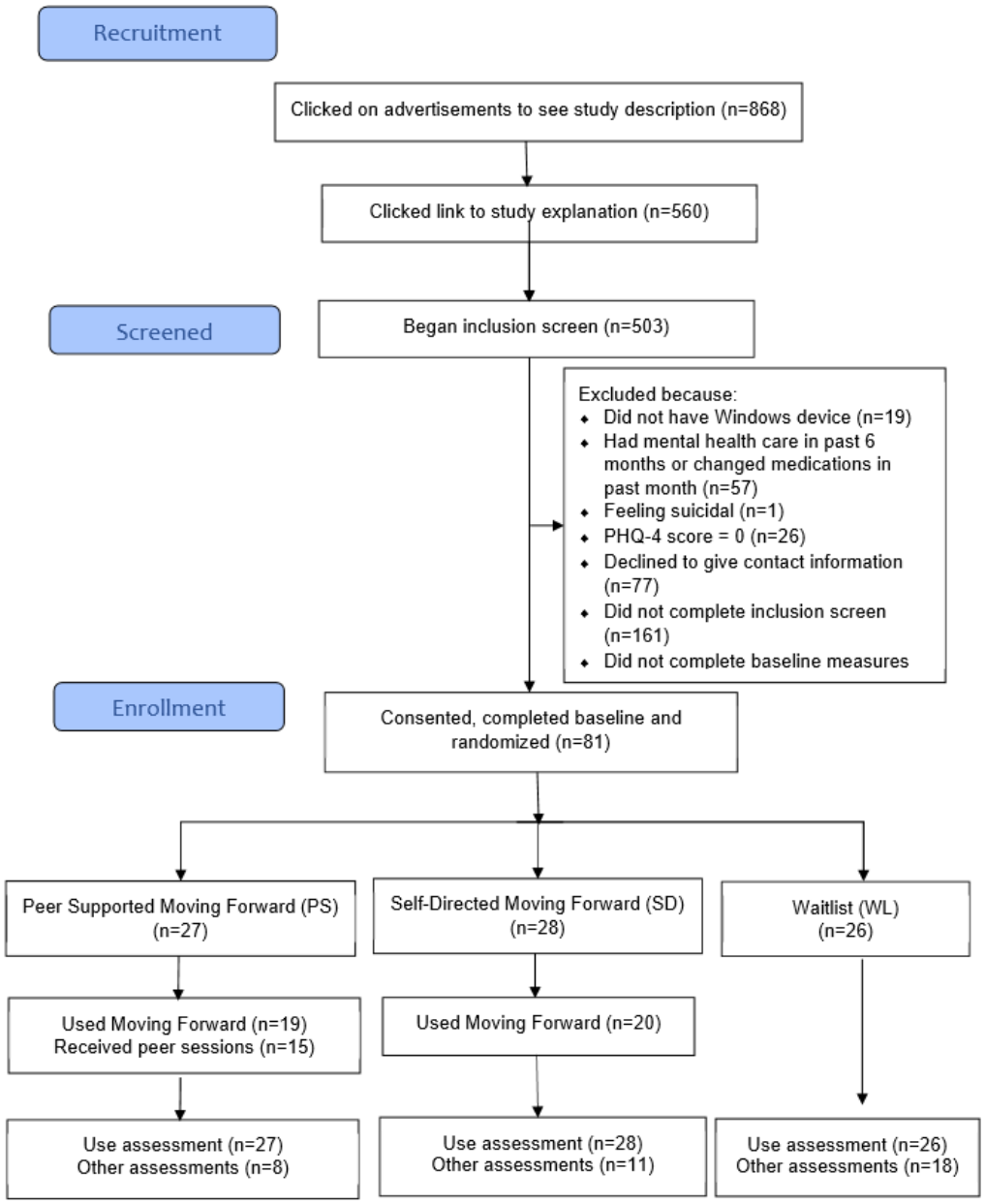
Veterans with unmet mental health needs were recruited for this study via Facebook advertisements. Figure 1 shows the flow of participants from having a Facebook advertisement show up on their Facebook feed to the collection of follow-up assessments. For this study, a veteran was determined to have unmet mental health needs if they endorsed mental health symptoms on assessment measures and denied engagement in current mental health treatment. To be included in the study, participants needed to (1) self-identify as veterans, (2) endorse any depression or anxiety item on the 4-item Patient Health Questionnaire [20] as occurring for several days or more in the past week, (3) have a phone and a computer with internet access, (4) be willing to be randomized to 1 of 3 conditions, and (5) work with a peer. Veterans were excluded if they had (1) an active suicide plan, (2) changes in psychoactive medications in the past month, or (3) were receiving mental health care at the time of enrollment. We confirmed veteran status using methods previously established by Kramer et al [21], which involve asking for details of service and contact information. Advertisements targeting veterans appeared in the Facebook feeds of users identified as having an interest in veterans. The advertisements were linked to a website that provided information about the study and allowed viewers to opt for a screening that was conducted via a web-based data collection tool, REDCap (Research Electronic Data Capture) [22,23]. The system recorded time stamps for when enrollments occurred to determine how often participants used the system outside of typical business hours within their home time zone.

Those who were eligible for the study completed baseline measures on the web, including assessment of sociodemographic characteristics, mental health symptoms, and problem-solving characteristics. Participants were randomized by the web-based platform to 1 of 3 conditions: SD MF, PS MF, and a waitlist (WL) control group. The WL group received no intervention and participated in both study assessments before being offered access to the MF course with or without PS. After their baseline assessment, SD and PS participants were directed to MF and asked to complete the 5 modules over the next 8 weeks. PS participants also received up to 5 phone sessions with a peer.

Course use data were collected using a commercial web-based learning platform. Collection was automated via programming, which initiated daily use reports to be sent from the learning platform to a REDCap file. To have use data identified with anonymous study IDs, the enrollment of each participant triggered a REDCap application programming interface to send the participant's study ID number to a server programmed to set up a new user account in the web-based learning system. In this way, use data were automatically and anonymously associated with participant IDs. Programming was used to link the data management system to a web-based learning platform to study course use. The data management system also supported the delivery of peer services.

Symptoms and problem-solving measures were re-administered on the web at the end of the study (8 weeks post baseline), along with a satisfaction measure. Participants received 1 to 3 automated reminder emails over a 1- to 2-week period (spaced 3-5 days apart) with a link to complete follow-up measures. Participants who did not complete the assessment in the SD and WL conditions and some participants in the PS condition received at least one reminder phone call over a 2-week period (25/81, 31% were left voicemails, 27/81, 33% spoke directly to research staff, and 11/81,13% could not be reached). Peers reminded PS participants to complete the assessment at the final peer session (when this occurred). After the final follow-up assessment, participants were given a choice between ending study participation and agreeing to a qualitative phone interview. Participants were compensated with gift cards and were paid US $5 for answering baseline measures, US $10 for completing the end of study measures, and US $15 for completing a qualitative interview (US $30 maximum).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flowchart. PHQ-4: Patient Health Questionnaire-4.



Measures

To make the assessments as brief as possible, a subset of items from 3 standard mental health screening tools was used. Item-level depression, anxiety, and PTSD symptom data from a prior study of 232 VHA primary care patients (Carlson et al, unpublished data, 2021) and from the first 17 participants of this study were analyzed using multivariate regression to select subsets of items to assess these symptoms. Depression was assessed with 5 of the original 9 Patient Health Questionnaire-9 [24] items (anhedonia, feeling down, trouble sleeping, feeling like a failure, and trouble concentrating), resulting in a 5-item scale with high internal consistency (α=.80). Anxiety was assessed using 4 of the original 7 General Anxiety Disorder-7 items [25] (feeling nervous, worry, trouble relaxing, and restlessness), resulting in a 4-item scale with high internal consistency (α=.88). PTSD was assessed using 9 of the 17 original items from the Screen for Posttraumatic Stress Symptoms [26,27] (avoiding thoughts, avoiding situations, nightmares, feeling shaky, irritability, surroundings feeling unreal, remembering awful things, upset upon reminders, lost track of what is happening around me, flashbacks, and things seeming unreal), resulting in a 9-item scale with high internal consistency (α=.88). Scores on the briefer measures correlated between .96 and .97 with scores on the full measures in the 2 data sets. Problem-solving knowledge was assessed using 6 items related to the information included in the course (eg, multiple choice questions on how stress can affect problem-solving, good strategies for problem solving, and how to evaluate possible solutions). Problem-solving confidence was assessed using 3 items inquiring about agreement with statements regarding best ways to solve problems, feeling confident about solving future problems, and having strong problem-solving skills, all answered on a 4-point scale ranging from *I don't agree* to *I agree a lot*. The Client Satisfaction Questionnaire comprised 8 items administered at the end of the

study and assessed satisfaction with the services they received without differentiating among recruitment, assessment, and intervention services [28].

Qualitative interviews were conducted post treatment to assess participant reactions to MF with and without PS. Results related to which aspects of PS were most useful or least useful and perceptions of working with a peer versus a traditional mental health provider have been included in this report.

The MF Course

MF is an instructionally sound, web-based self-help program [29]. It was built collaboratively by VHA and DoD as part of a coordinated public health initiative to address the unmet mental health needs of returning service members. To reduce any barriers to use, it is completely free, anonymous, and available to the public. Users do not have to register or provide any personal information to use the program. A version of the course is available on the web [29]. The MF web-based intervention was based on a 4-session problem-solving workshop that has shown efficacy in affecting key targets [30] and makes extensive use of video, animation, and interactive exercises to engage participants. Its design was informed by the principles of adult learning theory and closely follows best practices in instructional systems design [31]. It includes detailed explanations of the relationships among thoughts, behaviors, and moods. The program places a strong emphasis on completing problem-solving worksheets and other tasks between sessions and uses a variety of techniques to help users monitor and challenge unhelpful thoughts. MF is based on cognitive behavioral treatment principles; however, to reduce stigma, it is presented to veterans as an *educational and life coaching program* that teaches skills and tools to solve stressful problems and overcome obstacles. Videos show highly relatable, demographically diverse characters applying problem-solving principles to address problems that veterans commonly experience, such as financial difficulties and relationship distress. Table 1 includes the major components of each module. The modules are self-paced but typically take approximately 20 minutes each to complete.

Table 1. Moving Forward modules and peer sessions.

Moving Forward content	Peer sessions (fidelity elements)
__[a]	Session 1 • Help the veteran choose a problem to focus on • Ask the veteran to complete module 0
Module 0: What type of problem solver are you? • Common problem-solving challenges, interactive stress game, introduction to problem solving • Problem-solving attitudes and approaches—vignettes and explanations • Optimistic versus pessimistic problem-solving attitudes • Problem-solving approaches—thoughtful planner, quick-fixer, and avoider • Problem-solver questionnaire and feedback—identifying strengths and weaknesses and getting personal insight into how one handles stressful situations	Session 2 • Ask about module completion • Discuss veteran's problem-solving strengths and weaknesses • Ask the veteran to complete modules 1 to 2

Module 1: Solve problems when your brain is overloaded
• How our brains get overloaded and our limited ability to multitask
• How to *externalize, simplify, and visualize* to minimize *brain overload*
Module 2: Solve problems under stress
• Survey to measure stress level
• How stress affects your mind, body, and behavior (and problem solving)
Stop and Slow Down steps toolkit with videos (eg, relaxation exercises) and a game (Rocket Commander) to demonstrate steps
	Session 3
	• Ask about module completion
	• Discuss the following:
	  • Strategies to try when the veteran experiences *brain overload*
	  • How stress affects problem solving
	  • Any relaxation strategies veteran tried
	• Ask the veteran to complete module 3

Module 3: Solve problems step-by-step
• *Planful* problem solving to focus on the Think and Act steps of problem solving
• Interactive problem-solving worksheets—fillable exercise to define the problem, obstacles, and courses of action
• Videos on evaluating and selecting courses of action
	Sessions 4
	• Ask about module completion
	• Discuss the following:
	  • Implementation of the problem-solving action plan
	  • Overcoming barriers to the action plan
	• Create a new problem-solving worksheet, if applicable
	• Ask the veteran to keep working on their action plan and complete module 4

Module 4: Where to go from here
• Emphasizes practice and anticipating future problems; interactive game to illustrate this concept
• Encourages celebrating positive progress
• Encourages veterans to keep trying and offers vignette videos to demonstrate positive outcomes of using the program
• Offers examples of when and how to reach out for more help
	Session 5
	• Discuss the following:
	  • Implementation of the problem-solving action plan
	  • Overcoming barriers to the action plan
	• Review skills developed in the program
	• Develop a plan to continue problem solving without peer support
	• Connect to additional resources, as needed

[a]The participants began using Moving Forward after session 1 so no content is included in this cell.

Peer Support

The peer sessions were designed to be approximately 20-minute phone calls. Over 8 weeks, 5 sessions were offered to participants, and the spacing between sessions was based on participant preference. The session content was guided by the PS guide, which was adapted from a previous study of PS for a web-based mental health program [32]. The goals of the peer sessions were to help the participants fully engage in the MF content and apply it to daily life. Peers aimed to engage participants in meaningful discussions on how to apply problem-solving skills and share their own experiences of overcoming life's problems, as appropriate. Each session had essential fidelity elements, as detailed in Table 1. Peers used REDCap templates to document sessions and endorse the fidelity elements that were met. A total of 2 peers provided support for the study: both had experience working as peer specialists with veterans and completed study-specific training that included didactics, personally using MF, and receiving feedback on role-plays of sessions. Peers also participated in weekly group supervisions led by a clinical psychologist with peer supervision expertise.

Analyses

Feasibility and acceptability metrics were described using means and frequencies. Feasibility metrics included the rate of participant enrollment per month and the rates of assessment completion at baseline and at the end of the study. Peer fidelity was described as the percentage of fidelity elements completed. User acceptability metrics included the percentage of participants who logged in to use MF, percentage of participants who had any peer sessions, number of peer sessions completed, and participant satisfaction. Satisfaction among participants who did and did not have access to MF was compared using a

2-tailed *t* test. Poisson regression with a Pearson adjustment to the SE was used to compare the number of log-in days (a count variable) between the SD and PS conditions. Multilevel modeling (MLM) via SAS Proc Mixed was used to compare changes in problem-solving knowledge and problem-solving confidence among participants who used and did not use MF. MLM was also used to estimate the associations between problem-solving confidence and depression, anxiety, and PTSD symptom outcomes. MLM was also used to compare changes in mental health symptoms among participants who engaged in peer sessions with those who did not. Analyses were adequately powered for the primary (to test the feasibility and user-perceived acceptability of the methods) and secondary (to examine engagement in the web-based MF course for those with and without PS) aims. Our planned sample size did not provide sufficient power to detect differences between conditions in mental health symptoms; therefore, group-by-time analyses were not conducted.

Qualitative data were analyzed using a rapid analysis approach [33,34]. Templated summaries of interview responses were created and then entered into a matrix, with a domain name for each survey question placed on the horizontal axis and participants listed on the vertical axis. A total of 2 authors (JW and KP) independently began data reduction on discrete copies of the matrices to develop concise summaries that focused and organized the data. The authors then met to compare their data reductions and create a final matrix that reflected their agreed-upon themes.

Results

Sample Characteristics

A total of 81 veterans recruited from social media posts were enrolled and randomized to study conditions (PS 27/81, 33%; SD 28/81, 35%; WL 26/81, 32%). Participant characteristics are shown in Table 2. Participants lived in 38 different states, and most enrollments (60/81, 71%) occurred outside of typical office hours within the veterans' home time zone. The participants reported significant unmet mental health needs based on our definition in this study.

Table 2. Characteristics of the participants (N=81).

Variable	Participants
Age (years), mean (SD; range)	54 (9.4; 30-77)
Gender, n (%)	
Male	48 (59)
Race and ethnicity, n (%)	
White	60 (74)
Black	8 (10)
Latinx	8 (10)
Mixed race	8 (10)
Native American	8 (10)
Asian or Pacific Islander	3 (4)
Marital status, n (%)	
Single	7 (9)
Married or partnered	36 (44)
Separated or divorced	30 (37)
Widowed	4 (5)
Education, n (%)	
High school	11 (14)
Some college or vocational school	49 (60)
Bachelor's degree or more	31 (26)
Employment status, n (%)	
Employed	25 (31)
Disabled	22 (27)
Retired	16 (20)
Unemployed	11 (14)
Student or homemaker	4 (5)
Service branch, n (%)	
Army	96 (51)
Navy	42 (22)
Air Force	32 (17)
Marines	19 (10)
Coast Guard	1 (1)
Mental health, n (%)	
Endorsed all 5 symptoms of depression	35 (43)
Endorsed ≥3 symptoms of depression	67 (83)
Scored above clinical cutoff for depression on Patient Health Questionnaire-2 [35]	55 (68)
Endorsed ≥3 anxiety symptoms	66 (81)
Endorsed ≥6 posttraumatic stress disorder symptoms	53 (65)
Endorsed problems with sleep	74 (91)

Feasibility

Figure 1 shows the rates of participant recruitment, screening, enrollment, assessment, and intervention allocation. The web-based recruitment method yielded 13.5 participants per month over a 6-month period. Only 24% (19/81) of participants completed the follow-up assessment after receiving automated emails. The end of the study assessments were ultimately

completed by 46% (37/81) of the participants following research assistant phone calls. Assessment completion rates were 30% (8/27) for PS, 39% (11/28) for SD, and 69% (18/26) for WL conditions. Baseline characteristics (eg, gender, race, age, and depression, anxiety, or PTSD symptoms) did not significantly differ among participants who completed the follow-up assessment and those who did not across the entire sample or within randomized conditions.

Peer Sessions and Fidelity

Just over half (15/27, 56%) of the participants assigned to receive PS participated in sessions with peers. Of the 15 participants, 4 (15%) received 1 session, 7 (26%) received 2 to 4 sessions, and 4 (15%) received all 5 peer sessions. Moreover, 1 peer engaged and provided support sessions to 65% (12/18) of those assigned to work with him and had a mean of 2.9 (SD 1.6) sessions with the participants he did have sessions with, whereas a second peer provided sessions to 42% (3/7) of those assigned and had a mean of 2.2 (SD 1.4) sessions with those he did have sessions with. Peer sessions had 100% fidelity to the Peer Guide, indicating that the peer training methods combined with the REDCap documentation system were highly feasible for peers to complete.

User Acceptability

Most participants in the SD and PS conditions logged into the course at least once (19/27, 70% and 20/28, 71% of participants, respectively). Participant satisfaction was favorable in both the SD and PS conditions. Satisfaction was significantly higher in the SD and PS conditions compared with the WL condition (t_{35}=−2.54; *P*=.03) and did not differ between the SD and PS conditions. Specifically, 95% (18/19) of participants in the SD and PS conditions who completed the satisfaction measure rated the overall quality as good to excellent and said that they received the kind of services they wanted. Similarly, 90% (17/19) said that they would recommend the program to a friend.

Of the 8 PS participants who completed the follow-up interview, 4 (50%) agreed to the interviews, and 3 (38%) completed the interviews, all of whom reported highly positive feedback. Regarding their overall experience with the peer, participants noted that the peer they worked with was *extremely helpful*, encouraged them to continue in the course, and helped them feel like they were *not alone* in their struggles. The most helpful aspects of PS included having the peers clarify course content and suggest new ideas for problem solving when the participants felt stuck. A veteran noted, “having interaction with the website is one thing, but having someone to talk to and getting feedback is really helpful and supportive.” Participants commented that their peer’s veteran status made them feel like they had more in common with the peer and that the peer could understand their struggles more. A total of 2 suggestions for improvements were noted: 1 participant said that the 20-minute phone sessions felt “a bit rushed” and that having a picture of the peer would be helpful as they only spoke by phone.

Impact of PS on Course Use and Mental Health Symptoms

Table 3 shows the results of the analyses of course use. Course use was significantly greater for PS participants than for SD participants. The effect of PS was more pronounced among participants who engaged in at least one peer session; participants who had one or more PS sessions logged in on significantly more days than participants who had access to MF but had no PS sessions. Participants who had at least one PS session had larger decreases in depression than those who did not have a PS session. Participants who engaged in peer sessions did not have significantly different changes in anxiety symptoms compared with those who did not have peer sessions.

Table 3. Impact of peer support on course use and mental health symptoms (N=81).

Outcome	Values, n (%)	Values, mean (SE)	B (SE)	*P* value
Peer support				
Course use (days logged in)	27 (33)	2.96 (0.22)	−0.89 (0.41)	.03
Self−directed				
Course use (days logged in)	28 (35)	1.21 (0.67)	−0.89 (0.41)	.03
Any peer sessions				
Course use (days logged in)	15 (19)	4.87 (0.18)	1.56 (0.31)	.001
Depression symptoms change	15 (19)	−4.3 (1.93)	−6.73 (2.04)	.002
Anxiety symptoms change	15 (19)	−1.8 (2.17)	−3.88 (2.30)	.10
No peer sessions				
Course use (days logged in)	42 (52)	1.02 (0.25)	1.56 (0.31)	.001
Depression symptoms change	42 (52)	2.4 (0.68)	−6.73 (2.04)	.002
Anxiety symptoms change	42 (52)	2.1 (0.77)	−3.88 (2.30)	.10

Associations of Problem-Solving Confidence With MF Use and Mental Health Symptoms

Participants who used MF at least once (including those in the PS and MF conditions) had larger increases in problem-solving confidence compared with participants who did not use MF (2.4 vs −0.72; B=−1.72, SE 0.64; *P*=.01). Problem-solving knowledge did not differ by whether participants used MF (B=−0.35, SE 0.51; *P*=.31). In addition, a 1-point increase in problem-solving confidence was associated with a 1.12-point decrease in depression (B=−1.12, SE 0.20; *P*=.001), a 0.82-point decrease in anxiety (B=−0.82, SE 0.20; *P*=.001), and a 1.51-point decrease in PTSD severity (B=−1.51, SE 0.45; *P*=.001).

Discussion

Principal Findings

This paper reports on the feasibility and user-perceived acceptability of innovative and automated methods for recruiting, enrolling, assessing, and delivering a web-based problem-solving intervention (MF) to veterans with unmet mental health needs. The results also demonstrate the impact of PS on MF use and describe the relationships between multiple potential mechanisms and outcomes of web-based problem-solving training to inform future research.

The methods for social media recruitment were successful in recruiting a sample of veterans in the community with unmet mental health needs who were not engaged in mental health care in the previous 6 months. The sample was diverse in geographic location, age, gender, race, and ethnicity and resulted in the recruitment of many women and younger veterans who can be challenging to recruit with in-person recruitment methods in VHA. In addition, most participants enrolled in the evening or on a weekend day, indicating that they preferred to use the service outside of regular business hours.

Our automated systems for screening, enrollment, administration of the initial assessment, randomization, and condition assignment created a seamless, brief, and easy user experience that led to the enrollment of our targeted number of eligible participants. The automated collection of course use data gathered objective information about whether the course was accessed and how many days participants logged in to the course. The results indicated that the automated enrollment and assessment methods led to acceptable levels of eHealth engagement [36]. Our study found a similar or lower percentage of nonusers (16/55, 29%) compared with other eHealth studies, which found that 25% to 58% of the participants never used the intervention [37]. Clinical programs also often suffer from less-than-ideal engagement. For example, a report of a program found that less than half of primary care patients who screened positive for depression or PTSD and were offered mental health care received any sessions [4].

Given the widespread challenges of engaging individuals with mental health needs to use eHealth tools and mental health services more generally, our finding that PS was associated with increased MF use is significant. These results are supported by our qualitative findings, where participants described their peer relationships as encouraging and helpful in moving them through barriers in their own understanding of MF content or their ability to apply it to their current problems.

Our finding that 55% (15/27) of participants offered PS chose not to do phone sessions indicates that not all eHealth users want or need PS. It is possible that the delivery of PS by phone hampered uptake and that some users would have preferred messaging or video formats. An option for peer messaging or video chats may be feasible going forward, given the increase in remote delivery of services during the COVID-19 pandemic [7]. Satisfaction was also high in the SD MF condition, further indicating that many users did not see the need for additional support in using the modules. Although satisfaction results may be influenced by ceiling effects on the Client Satisfaction Questionnaire, taken together, these results signal that alternative study designs and stepped-care interventions may be more appropriate for studying the added benefits of PS to eHealth. For instance, a design that allows users to opt for PS if they want or need it or step up to additional support if they are not engaging in SD use could reveal how often users want and need additional support and if this additional support increases eHealth use among participants who cannot engage on their own.

The streamlined user experience created by the automated enrollment and assessment systems may have resulted in some veterans being enrolled and randomized who were not motivated to engage in follow-up assessments, as evidenced by our assessment attrition rate of 54%. High attrition from eHealth studies is common and well-described in the literature [16]. Efforts to boost follow-up assessments can include reducing participant burden (ie, number of questions) in assessments or adding interactions with research staff at enrollment via email, text, or phone to confirm that the participant truly intended to fully participate. Interaction with research staff can also build a participant's commitment to a study. A recent meta-analysis found that mobile health studies that enrolled participants on the web have much higher attrition (43%) than those that use in-person (11%) and phone (18%) enrollment [37]. Although adding contact with research staff is likely to decrease attrition, therefore increasing the internal validity of a study, it may also make the user experience more complex and less reflective of real-world eHealth use, decreasing the external validity of a study; that is, veterans who could benefit most from web-based programs may not meet a higher bar for motivation. Furthermore, adding contact with staff will increase study costs and decrease the number of people who can be reached. Ultimately, researchers may need to anticipate high assessment attrition in eHealth studies that use automated enrollment procedures and increase enrollment with the goal of still having an adequate sample size at the end of the study.

Our findings were consistent with the results of a meta-analysis that found higher assessment attrition in active conditions compared with WL controls and concluded that this might be because participants in active conditions have full access to the eHealth intervention and therefore are less motivated to complete assessments than those in the WL control who need to complete assessments before gaining access [37]. An additional explanation is that the active conditions (especially

PS in our study) required more participant effort and time than the WL control; therefore, WL participants may have been more willing to spend additional time on the assessments.

Confidence in problem-solving skills appears to be an important correlate of increased course use and changes in mental health symptoms. Confidence and the closely related construct of self-efficacy are common mechanisms of action in many behavioral interventions, including eHealth programs [14,38]. Future research should examine problem-solving confidence or mental health self-efficacy as a mediator between MF use and mental health outcomes and consider optimizing components of the intervention that increase confidence, such as peer encouragement to practice problem solving, as more practice is likely to lead to more mastery and confidence.

Engagement in PS was associated with decreases in depression. This is consistent with emerging research showing the mental health benefits of PS [39]. There are many possible mechanisms that may explain the relationship between PS and mental health outcomes that are well-described in the supportive accountability model, including that being accountable to a supportive and knowledgeable individual increases engagement with eHealth or self-help tools [17]. Peers add an additional component of shared experience that can enhance the strength and benefits of the relationship.

This study has several strengths and limitations. Automated methods were successfully used to recruit, screen for inclusion, obtain consent, and enroll veterans with unmet mental health needs to a study at rates comparable with other eHealth studies and referral in primary care. Baseline and course use data were also successfully collected using automated methods. Web-based recruitment resulted in a sample that was diverse in age, gender, service branch, service era, and geographic location. The study successfully randomized some participants to receive PS for the course and tracked remotely provided PS contacts and sessions. A major weakness of the study was the low rate of completion of the study end assessments. Automated requests to complete the assessments were not sufficiently effective to achieve adequate completion rates, and staff phone calls to encourage completion were inconsistent. To increase the study end measure completion in future studies, staff efforts to reach participants could be increased, and data might be collected by study staff by phone. Given the relatively low completion rate for the study end measures, it is possible that more satisfied participants completed measures at higher rates than less satisfied participants, possibly leading to a positive bias in both quantitative and qualitative satisfaction results. An additional limitation was that the success in engaging participants was lower for one of the peer specialists. Future studies on the impact of PS on eHealth use should consider increasing the number of PS providers and assessing factors such as participants' perceptions of therapeutic alliance with peers.

Conclusions

In conclusion, automated processes for recruiting, enrolling, screening, assessing, and providing a cognitive behavioral eHealth intervention are feasible and acceptable overall; however, additional efforts are necessary to achieve adequate study end assessment completion rates. The finding that delivering PS for MF by phone with high fidelity was feasible and the increased use of the course indicates that PS may boost engagement with web-based courses. Innovative designs such as SMARTs (Sequential Multiple Assignment Randomization Trials) may clarify for whom PS is most helpful, when it is most helpful, and what outcomes are likely to improve. Finally, the associations of problem-solving confidence with both course use and improved mental health outcomes indicate that problem-solving confidence shows promise as a potential mechanism of action in future eHealth research.

Acknowledgments

This paper is based on work supported by the Department of Veterans Affairs, Veterans Health Administrative Office of Research and Development, Health Services Research and Development, HX002361-02.

The views expressed in this paper are those of the authors and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the US government.

Conflicts of Interest

None declared.

References

1. Results from the 2016 national survey on drug use and health: detailed tables. Substance Abuse and Mental Health Services Administration Center for Behavioral Health Statistics and Quality. 2017. URL: https://www.samhsa.gov/data/sites/default/files/NSDUH-DetTabs-2016/NSDUH-DetTabs-2016.pdf [accessed 2021-11-27]
2. Hoge CW, Grossman SH, Auchterlonie JL, Riviere LA, Milliken CS, Wilk JE. PTSD treatment for soldiers after combat deployment: low utilization of mental health care and reasons for dropout. Psychiatr Serv 2014 Aug 01;65(8):997-1004. [doi: 10.1176/appi.ps.201300307] [Medline: 24788253]
3. Aakre JM, Himelhoch S, Slade EP. Mental health service utilization by Iraq and Afghanistan veterans after entry into PTSD specialty treatment. Psychiatr Serv 2014 Aug 01;65(8):1066-1069. [doi: 10.1176/appi.ps.201300117] [Medline: 24882427]

4. Lindley S, Cacciapaglia H, Noronha D, Carlson E, Schatzberg A. Monitoring mental health treatment acceptance and initial treatment adherence in veterans: veterans of Operations Enduring Freedom and Iraqi Freedom versus other veterans of other eras. Ann N Y Acad Sci 2010 Oct;1208:104-113. [doi: 10.1111/j.1749-6632.2010.05692.x] [Medline: 20955332]
5. Fischer E, McSweeney J, Wright P, Cheney A, Curran G, Henderson K, et al. Overcoming barriers to sustained engagement in mental health care: perspectives of rural veterans and providers. J Rural Health 2016 Sep;32(4):429-438. [doi: 10.1111/jrh.12203] [Medline: 27557334]
6. Possemato K, Wray LO, Johnson E, Webster B, Beehler GP. Facilitators and barriers to seeking mental health care among primary care veterans with posttraumatic stress disorder. J Trauma Stress 2018 Oct;31(5):742-752. [doi: 10.1002/jts.22327] [Medline: 30338576]
7. Rosen CS, Morland LA, Glassman LH, Marx BP, Weaver K, Smith CA, et al. Virtual mental health care in the Veterans Health Administration's immediate response to coronavirus disease-19. Am Psychol 2021 Jan;76(1):26-38. [doi: 10.1037/amp0000751] [Medline: 33119331]
8. Erbes CR, Stinson R, Kuhn E, Polusny M, Urban J, Hoffman J, et al. Access, utilization, and interest in mHealth applications among veterans receiving outpatient care for PTSD. Mil Med 2014 Nov;179(11):1218-1222. [doi: 10.7205/MILMED-D-14-00014] [Medline: 25373044]
9. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
10. Fortney JC, Burgess JF, Bosworth HB, Booth BM, Kaboli PJ. A re-conceptualization of access for 21st century healthcare. J Gen Intern Med 2011 Nov;26 Suppl 2:639-647 [FREE Full text] [doi: 10.1007/s11606-011-1806-6] [Medline: 21989616]
11. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
12. Sander L, Rausch L, Baumeister H. Effectiveness of internet-based interventions for the prevention of mental disorders: a systematic review and meta-analysis. JMIR Ment Health 2016 Aug 17;3(3):e38 [FREE Full text] [doi: 10.2196/mental.6061] [Medline: 27535468]
13. Funderburk JS, Dollar KM, King PR, Pomerantz AS, Bernstein L, Aspnes A, et al. Problem-solving training for VA integrated primary care providers: real-world outcomes. Professional Psychol Res Practice 2021 Feb;52(1):58-66. [doi: 10.1037/pro0000340]
14. Clarke J, Proudfoot J, Birch M, Whitton AE, Parker G, Manicavasagar V, et al. Effects of mental health self-efficacy on outcomes of a mobile phone and web intervention for mild-to-moderate depression, anxiety and stress: secondary analysis of a randomised controlled trial. BMC Psychiatry 2014 Sep 26;14:272 [FREE Full text] [doi: 10.1186/s12888-014-0272-1] [Medline: 25252853]
15. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
16. Eysenbach G. The law of attrition. J Med Internet Res 2005 Mar 31;7(1):e11 [FREE Full text] [doi: 10.2196/jmir.7.1.e11] [Medline: 15829473]
17. Mohr DC, Cuijpers P, Lehman K. Supportive accountability: a model for providing human support to enhance adherence to eHealth interventions. J Med Internet Res 2011 Mar 10;13(1):e30 [FREE Full text] [doi: 10.2196/jmir.1602] [Medline: 21393123]
18. Fortuna KL, Naslund JA, LaCroix JM, Bianco CL, Brooks JM, Zisman-Ilani Y, et al. Digital peer support mental health interventions for people with a lived experience of a serious mental illness: systematic review. JMIR Ment Health 2020 Apr 03;7(4):e16460 [FREE Full text] [doi: 10.2196/16460] [Medline: 32243256]
19. Peer specialist toolkit: implementing peer support services in VHA. U.S. Department of Veterans Affairs. URL: https://www.mirecc.va.gov/visn4/peer_specialist_toolkit.asp [accessed 2021-11-27]
20. Kroenke K, Spitzer RL, Williams JB, Löwe B. An ultra-brief screening scale for anxiety and depression: the PHQ-4. Psychosomatics 2009;50(6):613-621. [doi: 10.1176/appi.psy.50.6.613] [Medline: 19996233]
21. Kramer J, Rubin A, Coster W, Helmuth E, Hermos J, Rosenbloom D, et al. Strategies to address participant misrepresentation for eligibility in Web-based research. Int J Methods Psychiatr Res 2014 Mar;23(1):120-129 [FREE Full text] [doi: 10.1002/mpr.1415] [Medline: 24431134]
22. Harris PA, Taylor R, Minor BL, Elliott V, Fernandez M, O'Neal L, REDCap Consortium. The REDCap consortium: building an international community of software platform partners. J Biomed Inform 2019 Jul;95:103208 [FREE Full text] [doi: 10.1016/j.jbi.2019.103208] [Medline: 31078660]
23. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)--a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009 Apr;42(2):377-381 [FREE Full text] [doi: 10.1016/j.jbi.2008.08.010] [Medline: 18929686]
24. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]

25. Spitzer RL, Kroenke K, Williams JB, Löwe B. A brief measure for assessing generalized anxiety disorder: the GAD-7. Arch Intern Med 2006 May 22;166(10):1092-1097. [doi: 10.1001/archinte.166.10.1092] [Medline: 16717171]
26. Carlson EB. Psychometric study of a brief screen for PTSD: assessing the impact of multiple traumatic events. Assessment 2001 Dec;8(4):431-441. [doi: 10.1177/107319110100800408] [Medline: 11785587]
27. Caspi Y, Carlson EB, Klein E. Validation of a screening instrument for posttraumatic stress disorder in a community sample of Bedouin men serving in the Israeli Defense Forces. J Trauma Stress 2007 Aug;20(4):517-527. [doi: 10.1002/jts.20237] [Medline: 17721959]
28. Attkisson C, Zwick R. The client satisfaction questionnaire. Eval Program Plan 1982 Jan;5(3):233-237. [doi: 10.1016/0149-7189(82)90074-x]
29. Moving forward: overcome life's challenges. U.S. Department of Veterans Affairs. URL: https://www.veterantraining.va.gov/movingforward/ [accessed 2021-11-27]
30. Tenhula WN, Nezu AM, Nezu CM, Stewart MO, Miller SA, Steele J, et al. Moving forward: a problem-solving training program to foster veteran resilience. Prof Psychol Res Pract 2014;45(6):416-424. [doi: 10.1037/a0037150]
31. Dick W, Carey L. The Systematic Design of Instruction. New York City, New York: Pearson, 6th Edition; 2004.
32. Possemato K, Johnson EM, Emery JB, Wade M, Acosta MC, Marsch LA, et al. A pilot study comparing peer supported web-based CBT to self-managed web CBT for primary care veterans with PTSD and hazardous alcohol use. Psychiatr Rehabil J 2019 Sep;42(3):305-313 [FREE Full text] [doi: 10.1037/prj0000334] [Medline: 30489140]
33. Gale RC, Wu J, Erhardt T, Bounthavong M, Reardon CM, Damschroder LJ, et al. Comparison of rapid vs in-depth qualitative analytic methods from a process evaluation of academic detailing in the Veterans Health Administration. Implement Sci 2019 Feb 01;14(1):11 [FREE Full text] [doi: 10.1186/s13012-019-0853-y] [Medline: 30709368]
34. Hamilton A. Qualitative methods in rapid turn-around health services research. U.S. Department of Veterans Affairs. 2013. URL: https://www.hsrd.research.va.gov/for_researchers/cyber_seminars/archives/video_archive.cfm?SessionID=780 [accessed 2021-11-27]
35. Löwe B, Kroenke K, Gräfe K. Detecting and monitoring depression with a two-item questionnaire (PHQ-2). J Psychosom Res 2005 Feb;58(2):163-171. [doi: 10.1016/j.jpsychores.2004.09.006] [Medline: 15820844]
36. Wangberg SC, Bergmo TS, Johnsen JK. Adherence in internet-based interventions. Patient Prefer Adherence 2008 Feb 02;2:57-65 [FREE Full text] [Medline: 19920945]
37. Linardon J, Fuller-Tyszkiewicz M. Attrition and adherence in smartphone-delivered interventions for mental health problems: a systematic and meta-analytic review. J Consult Clin Psychol 2020 Jan;88(1):1-13. [doi: 10.1037/ccp0000459] [Medline: 31697093]
38. Sheeran P, Maki A, Montanaro E, Avishai-Yitshak A, Bryan A, Klein WM, et al. The impact of changing attitudes, norms, and self-efficacy on health-related intentions and behavior: a meta-analysis. Health Psychol 2016 Nov;35(11):1178-1188. [doi: 10.1037/hea0000387] [Medline: 27280365]
39. Chinman M, George P, Dougherty RH, Daniels AS, Ghose SS, Swift A, et al. Peer support services for individuals with serious mental illnesses: assessing the evidence. Psychiatr Serv 2014 Apr 01;65(4):429-441. [doi: 10.1176/appi.ps.201300244] [Medline: 24549400]

Abbreviations

DoD: Department of Defense
MF: Moving Forward
MLM: multilevel modeling
PS: peer support
PTSD: posttraumatic stress disorder
REDCap: Research Electronic Data Capture
SD: self-directed
SMART: Sequential Multiple Assignment Randomization Trial
VHA: Veterans Health Administration
WL: waitlist

Edited by R Kukafka; submitted 12.04.21; peer-reviewed by K Lehman, L Martinengo, CM Schieszler-Ockrassa; comments to author 22.05.21; revised version received 13.08.21; accepted 14.10.21; published 13.01.22.

Please cite as:
Possemato K, Wu J, Greene C, MacQueen R, Blonigen D, Wade M, Owen J, Keane T, Brief D, Lindley S, Prins A, Mackintosh MA, Carlson E
Web-Based Problem-solving Training With and Without Peer Support in Veterans With Unmet Mental Health Needs: Pilot Study of Feasibility, User Acceptability, and Participant Engagement
J Med Internet Res 2022;24(1):e29559
URL: https://www.jmir.org/2022/1/e29559
doi:10.2196/29559
PMID:35023846

Original Paper

The Relative Contributions of Live and Recorded Online Mindfulness Training Programs to Lower Stress in the Workplace: Longitudinal Observational Study

Ruth Q Wolever[1], PhD; Michael T M Finn[2], PhD; Denise Shields[3], PhD

[1]Osher Center for Integrative Health at Vanderbilt, Department of Physical Medicine & Rehabilitation, Vanderbilt University School of Medicine, Nashville, TN, United States

[2]Department of Pediatrics & Human Development, Michigan State University, Grand Rapids, MI, United States

[3]eMindful Inc, Orlando, FL, United States

Corresponding Author:
Ruth Q Wolever, PhD
Osher Center for Integrative Health at Vanderbilt
Department of Physical Medicine & Rehabilitation
Vanderbilt University School of Medicine
3401 West End Ave
Suite 380
Nashville, TN, 37203
United States
Phone: 1 919 622 2819
Email: ruth.wolever@vumc.org

Abstract

Background: Despite numerous gaps in the literature, mindfulness training in the workplace is rapidly proliferating. Many "online" or "digital mindfulness" programs do not distinguish between live teaching and recorded or asynchronous sessions, yet differences in delivery mode (eg, face-to-face, online live, online self-guided, other) may explain outcomes.

Objective: The aim of this study was to use existing data from an online mindfulness solutions company to assess the relative contribution of live and recorded mindfulness training to lower perceived stress in employees.

Methods: Perceived stress and the amount of live and recorded online mindfulness training accessed by employees were assessed during eMindful's One-Percent Challenge (OPC). The OPC is a 30-day program wherein participants are encouraged to spend 1% of their day (14 minutes) practicing mindfulness meditation on the platform. We used linear mixed-effects models to assess the relationship between stress reduction and usage of components of the eMindful platform (live teaching and recorded options) while controlling for potential reporting bias (completion) and sampling bias.

Results: A total of 8341 participants from 44 companies registered for the OPC, with 7757 (93.00%) completing stress assessments prior to the OPC and 2360 (28.29%) completing the postassessment. Approximately one-quarter of the participants (28.86%, 2407/8341) completed both assessments. Most of the completers (2161/2407, 89.78%) engaged in the platform at least once. Among all participants (N=8341), 8.78% (n=707) accessed only recorded sessions and 33.78% (n=2818) participated only in the live programs. Most participants engaged in both live and recorded options, with those who used any recordings (2686/8341, 32.20%) tending to use them 3-4 times. Controlling for completer status, *any* participation with the eMindful OPC reduced stress (B=–0.32, 95% CI –0.35 to –0.30, SE=0.01, $t_{2393.25}$=–24.99, P<.001, Cohen d=–1.02). Participation in live programs drove the decrease in stress (B=–0.03, SE=0.01, $t_{3258.61}$=–3.03, P=.002, d=–0.11), whereas participation in recorded classes alone did not. Regular practice across the month led to a greater reduction in stress.

Conclusions: Our findings are in stark contrast to the rapid evolution of online mindfulness training for the workplace. While the market is reproducing apps and recorded teaching at an unprecedented pace, our results demonstrate that live mindfulness programs with recorded or on-demand programs used to supplement live practices confer the strongest likelihood of achieving a significant decrease in stress levels.

(J Med Internet Res 2022;24(1):e31935) doi:10.2196/31935

KEYWORDS

mindfulness; live versus recorded; synchronous; asynchronous; workplace; digital health; online; live teaching; contribution; training; stress; longitudinal; observational; platform; eHealth; mental health

Introduction

As the practice of mindfulness continues to rise [1,2], there has been a concomitant emergence of online mindfulness programs [3]. Prior to the COVID-19 pandemic, an estimated 60% of mid-to-large–sized companies in the United States were offering mind-body programs such as mindfulness for their employees [4], with related services creating a billion-dollar industry [5]. The pandemic-related spike in stress and mental health issues [6,7] and the fact that such spikes impair multiple aspects of adult learning [8,9] have led to an even greater need to support employees [10,11]. As rapid growth of the online and workplace mindfulness industries has outpaced research, there is an increased demand for rigorous methodology to carefully evaluate workplace mindfulness training, particularly online training. Relevant literature has multiple confounds, including the fact that the relative impact of live (synchronous) and recorded (asynchronous) digital mindfulness training for the workplace is completely unknown.

Multiple systematic reviews have demonstrated solid evidence that mindfulness training has a positive impact on stress and other indicators of mental health [12-16]. In addition, mindfulness interventions clearly reduce anxiety and depression symptoms in a therapeutic context [17,18], and help even as a stand-alone intervention [19]. Mindfulness programs have consistently been shown to assist clinical patients, health care professionals, and multiple types of students, and further promise benefits for the workplace [20-22] by helping individuals to develop several skills [23]. For example, mindfulness meditation improves generalized attention, alerting, and multiple aspects of executive control [24,25]. There is a solid rationale for why mindfulness training may enhance workplace functioning [20]. Nonreactive attention on the present moment should enhance focus while lowering errors; practicing curiosity and new perspectives should invite creative problem-solving; and cultivation of the attitudes of mindfulness should enhance interpersonal relationships, customer service, and leadership [20]. Although little empirical work has actually assessed the potential improvement in occupational functioning [26], very recent meta-analyses on mindfulness interventions in the workplace show robust and consistent improvements on stress and indicators of well-being. One recent meta-analysis of 56 randomized controlled trials (RCTs) found that, compared to controls, *live* workplace mindfulness programs did effectively lower stress, mental distress, and burnout, while enhancing well-being, mindfulness, compassion, and job satisfaction (Hedges g=0.32-0.77). Importantly, improvements were sustained at short-term follow-up assessments up to 12 weeks [27]. Similarly, another recent meta-analysis of 35 RCTs found medium effects on stress, distress, anxiety, depression, and burnout, while also revealing small to medium effects for health, job performance, compassion/empathy, and well-being [21]. A third recent meta-analysis of 23 RCTs concurred in finding moderate effects for stress, distress, anxiety, and well-being [22]. Fortunately, this study further delved into the heterogeneity and rigor of the studies. Analyses of this heterogeneity and risk of bias clarified that findings for burnout, depression, and work performance measures are premature. Study method variability included differences in measures used, assessment timelines, selection and attrition bias, and limited use of active control groups. Intervention variability factors included differences in training content, dose (program hours, weeks), and delivery mode (eg, face to face, online live, online self-guided, other) [22]. Other researchers have also identified concerns with the state of the literature on mindfulness training in the workplace, noting its promise yet need to uplevel the rigor of studies to manage threats to internal and external validity [26]. For example, workplace-based mindfulness programs are typically adapted from training programs in other contexts, but do not generally utilize the same training structure, protocol, or time commitment as the original studies [22]. Although these adaptations may fit well in busy workplace settings, workplace programs are often assumed to be supported by evidence from the original studies while actually lacking empirical validation. Moreover, the literature on mindfulness at the worksite has largely relied upon cross-sectional studies rather than tracking longitudinal outcomes such as performance over time [28].

These literature shortcomings leave large gaps in understanding what moderates effective outcomes for mindfulness programs in the workplace. One significant moderator may well be program delivery venues. The same mindfulness training program will produce similar outcomes whether delivered in person or online, provided it is delivered live (ie, in real time or synchronous) [16]. Meta-analyses of 8 RCTs demonstrated a medium effect size (Hedges g=0.432) for perceived stress and a small effect size (Hedges g=0.275) for mindfulness measures in nonclinical populations in fully online, *live* mindfulness programs [29]. However, there may be a difference in the impact of workplace mindfulness programs when provided live (whether in person or online) versus fully recorded (“on demand” or “asynchronous”) or delivered through static apps. Only one review assessed delivery methods and found no moderating effects for program characteristics [27]. However, the authors acknowledged that their study may have been underpowered to detect program differences. In addition, their included studies had an unequal distribution of mindfulness training delivery venues (eg, on site, online). Although the unequal distribution allowed for greater external validity, it may have skewed the findings. The authors suggested that further studies are needed to explore the potential differential effects of mindfulness program characteristics.

“Online” or “digital mindfulness” studies frequently do not distinguish between live and recorded training. Furthermore, app-based, recorded web-based, or live online teaching [3,30] are often not differentiated, despite the dearth of evidence for all but live, online teaching. Researchers note the substantial amount of human and financial resources required to provide participants with live, tailored feedback versus automated

feedback [29]. Given the relatively low cost and ease of reproducing automated feedback and on-demand content and apps, companies tend to "check the box" of providing mindfulness programs without empirical assessment of the program's impact or the methodology of delivery. As a result, many "on-demand" programs and apps are being offered to employers and consumers, inappropriately leveraging empirical research conducted on live mindfulness training programs. For example, there are more than 2500 mindfulness-based apps available [31], yet rigorous efficacy research for such apps is meager [30,32,33]. In fact, in a 2015 review and evaluation of 560 mindfulness apps, there was only one efficacy study [33]. In the last 5 years, at least 34 RCTs have examined a mindfulness intervention where the main component was app-based mindfulness practice [34]. Although meta-analyses showed significant improvements in stress (Hedges g=0.46) using 15 of the trials, when one outlier trial was excluded, the benefit dropped by 30% (Hedges g=0.32 for 14 trials). The impact on anxiety, depression, and psychological well-being was also significant, albeit with small effect sizes (Hedges g=0.28-0.33). No significant effects were seen for distress or general well-being in the studies of mindfulness apps. More concerning were the methodological issues noted, including the fact that only 12 of the 34 trials had a low risk of bias in 5 or 6 potential bias domains. In general, the studies of apps lacked information on randomization and concealment allocation, and less than a third reported intention-to-treat analyses. It is thus unclear how effective recorded practices and apps are when not supplemented by live teaching.

Online mindfulness programs for the workplace come in many forms, ranging from live programs using fully validated protocols to workplace adaptations with no empirical backing to recorded programs and apps. Importantly, the relative impact of live online teaching versus recorded training is completely unknown. Given the rapid proliferation of digital mindfulness training opportunities to help participants acquire the skills necessary to negotiate today's intense uncertainty and ongoing rise in stress, we wondered about the relative impact of participants' use of recorded training materials versus live training. The objective of this study was to use existing data from a mindfulness solutions company to assess the contribution of live and recorded training to lower perceived stress.

Methods

Study Design and Ethics

Anonymized data were used to assess the immediate longitudinal impact of participating in a 30-day online mindfulness training platform. We evaluated the contributions of participant use of live programs and recorded offerings in explaining potential change in perceived stress. This nonhuman research study was provided an exemption by the Vanderbilt University Medical Center Institutional Review Board.

Participants

Participants were unique users of eMindful Inc, who registered for the One Percent Challenge (OPC) between January 1, 2017, and February 29, 2020. Sociodemographic data on participants were obtained in two ways. Participants from the 2020 OPCs were asked to provide age, gender, and race/ethnicity in a preparticipation survey, but there was no requirement to do so. A second attempt at obtaining sociodemographic data was made by eMindful in culling the eligibility file data for earlier cohorts; data availability depended upon how the organization onboarded their employees with eMindful. In both cases, preferences of the individual organizations whose employees were participating were honored and data collection was often seen as a barrier to their employees getting started.

Mindfulness Program

The OPC is a 30-day program wherein participants are encouraged to spend 1% of their day (14 minutes) practicing mindfulness meditation on the eMindful platform. Sixteen live, guided trainings are offered each workday by highly skilled mindfulness teachers with an average of 21 years of personal practice, 76% of whom have advanced degrees (eg, PhD, MD) or are licensed clinicians and 70% of whom are certified mindfulness teachers from the International Mindfulness Teacher Association. The live mindfulness sessions come in 3 formats, with an average attendance of 39.3 participants per session. Attendance at a given session has a wide range (9 to 183 participants); there is very high participation in the weekday morning sessions and very low attendance late nights and weekends. The most common format is the 14-minute Mindful Daily (topic), which consists of 3-4 minutes of didactic teaching on the application of mindfulness to a particular topic (eg, trouble falling sleep, coworker conflict, managing your "to do" list), followed by 10 minutes of guided mindfulness practice and a closing inquiry. Additional Mindful Daily (practice) sessions are 15 and 30 minutes in length and consist only of guided practice. Finally, there are multisession live programs that are related to chronic conditions. The multisession programs are delivered in sequenced, 55-minute sessions that aim to build skills in a particular order. They also combine didactic teaching and experiential practice, but allow more time for personal inquiry. While the focus of eMindul has been live teaching in a virtual classroom, the company developed an app with over 2500 recorded sessions that are "on demand" for participants to use to supplement or replace the live programs. The "on demand" options include high-quality reproductions of the live Mindful Daily sessions as well as recorded practice sequences designed specifically for common concerns (eg, Seven Days of Managing Workplace Stress, Return to the Workplace, Gratefully all in). Companies who provide the OPC to their employees typically use the OPC to introduce the new benefit to employees. Promotional material provided to companies to encourage participation includes emails, one-page flyers, posters, and brief newsletters focused on the relevance of mindfulness for employee well-being.

Measures

Usage of the eMindful Platform

Backend data from eMindful provided two types of usage information: attendance (number of sessions attended, whether live or recorded) and days of mindfulness (number of days on which a participant practiced through use of the platform, no matter how many practices per day they completed).

Perceived Stress

Participants were asked to complete the Perceived Stress Scale-4 (PSS-4) [35] within 1 week of the start of the OPC and immediately post-OPC. Participants who did not respond postprogram were asked again through automated email and were allowed to take the PSS-4 up to 30 days after the closing of the OPC, regardless of the number of sessions completed. The PSS-4 is a 4-item abbreviated and validated version of the Perceived Stress Scale, which is a widely used measure of transactional stress that takes into consideration the balance between stressors and one's perceived ability to manage them. The scale asks the following questions: In the last month, how often have you felt that you were unable to control the important things in your life? In the last month, how often have you felt confident about your ability to handle your personal problems? In the last month, how often have you felt that things were going your way? In the last month, how often have you felt difficulties were piling up so high that you could not overcome them?

Items are scored on a 0-4 Likert scale of how often they were experienced in the past month (Never to Very Often). Two of the items are then reverse-scored and item scores are summed to provide a total score. In addition to calculating pre and post perceived stress levels, an absolute difference was calculated for each individual to provide estimates of change over time.

Statistical Plan and Data Presentation

Descriptive statistics were calculated on the sociodemographic, platform usage, and outcome data. We used linear mixed-effects models to assess the relationship between usage of the eMindful platform and reduction in stress. We were also interested in the relative impact of participation in live versus recorded sessions.

Given the practical nature of this real-world trial, and the level of potential nonrandom missingness, we first examined potential bias in completing both surveys on the basis of pre-OPC stress levels. This informed our inclusion of this factor in the modeling. We also had incomplete sociodemographic data that were assessed for potential sampling bias, and we examined the additional impact of relevant participant demographics on the model.

Results

Participants

A total of 8341 participants registered for the OPC. They were from 44 organizations who provided the OPC to their employees in the given timeline. Eight distinct OPC programs were run across the time period. The average number of participants per organization was 189.6 (SD 549.7), with a very wide range (1-3354 per organization). Thirty participants had a unique OPC that lasted 61 rather than 30 days, and hence were removed from the total count and analyses to avoid confounding the outcomes.

Participants who completed the preparticipation survey with sociodemographic data (age, gender, and race/ethnicity) or whose sociodemographic data were obtained from eligibility files allowed us to describe the sample. These two procedures produced information resulting in age being calculated for 68.04% (n=5675) of the total sample (N=8341), with a mean age of approximately 45 years, ranging from 20 to 79 years. Self-reported data on gender was available for 48.09% (n=4011) of the entire sample; the majority identified as female. Self-reported race/ethnicity data were available for 12.42% (n=1036) of the entire sample, with the majority identifying as white, followed by "nonwhite," and preferring not to answer. Baseline demographics are presented in Table 1 organized by completer status.

Table 1. Demographics by completer status.

Characteristic	Did not complete both assessments (n=5934)	Completers of pre and post PSS-4[a] (n=2407)	Total (N=8341)	*P* value
Age (years)				<.001
Responses, n	3509	2166	5675	
Mean (SD)	44.3 (11.02)	46.3 (10.91)	45.1 (11.02)	
Median (range)	44 (20-73)	47 (21-79)	45 (20-79)	
Gender				.09[b]
Female, n (%)	1871 (81.24)	1413 (82.73)	3284 (81.87)	
Male, n (%)	430 (18.67)	282 (16.51)	712 (17.75)	
Prefer not to answer, n (%)	2 (0.10)	13 (0.76)	15 (0.37)	
Missing, n	3631	699	4330	
Race				.67
White, n (%)	132 (65.7)	560 (67.1)	692 (66.8)	
Nonwhite, n (%)	56 (27.9)	220 (26.3)	276 (26.6)	
Prefer not to answer, n (%)	13 (6.5)	55 (6.6)	68 (6.6)	
Missing, n	5733	1572	7305	

[a]PSS-4: Perceived Stress Scale-4.

[b]Comparison based on male versus female.

Data Availability and Statistics

Usage of the eMindful Platform

An estimated 65.93% (n=5499) of the 8341 participants who registered for an OPC engaged in a mindfulness session at least once during an OPC. A greater proportion of participants accessed only live sessions compared with those who accessed only recorded sessions (Table 2).

Table 2. Platform usage by completer status.

Usage metric	Did not complete both assessments (n=5934)	Completed pre and post PSS-4[a] (n=2407)	Total (N=8341)	*P* value
Days (out of 30) used eMindful platform				<.001
N	5934	2407	8341	
Mean (SD)	3.0 (5.6)	14.7 (10.7)	6.3 (9.1)	
Median (range)	1.0 (0.0-30.0)	14.0 (0.0-30.0)	1.0 (0.0-30.0)	
Number of live programs				<.001
N	5934	2407	8341	
Mean (SD)	2.7 (6.0)	14.9 (14.8)	6.2 (10.9)	
Median (range)	0.0 (0.0-127.0)	12.0 (0.0-151.0)	1.0 (0.0-151.0)	
Days (out of 30) of live programs				<.001
N	5934	2407	8341	
Mean (SD)	2.2 (4.6)	11.9 (10.2)	5.0 (8.0)	
Median (range)	0.0 (0.0-30.0)	10.0 (0.0-30.0)	1.0 (0.0-30.0)	
Number of recorded programs				<.001
N	5934	2407	8341	
Mean (SD)	1.4 (4.8)	5.1 (10.9)	2.5 (7.3)	
Median (range)	0.0 (0.0-139.0)	0.0 (0.0-122.0)	0.0 (0.0-139.0)	
Days (out of 30) of recorded programs				<.001
N	5934	2407	8341	
Mean (SD)	0.9 (3.1)	3.4 (7.0)	1.7 (4.7)	
Median	0.0 (0.0-30.0)	0.0 (0.0-30.0)	0.0 (0.0-30.0)	
Any eMindful use at all, n (%)				<.001
No	2588 (43.61)	246 (10.22)	2834 (33.98)	
Yes	3346 (56.39)	2161 (89.78)	5507 (66.02)	
Only used live programs, n (%)				<.001
No	4112 (69.30)	1411 (58.62)	5523 (66.22)	
Yes	1822 (30.70)	996 (41.38)	2818 (33.78)	
Only used recorded programs, n (%)				<.001
No	5345 (90.07)	2289 (95.10)	7634 (91.52)	
Yes	589 (9.93)	118 (4.90)	707 (8.48)	
Used both live and recorded programs, n (%)				<.001
No	4999 (84.24)	1360 (56.50)	6359 (76.24)	
Yes	935 (15.76)	1047 (43.50)	1982 (23.76)	

[a]PSS-4: Perceived Stress Scale-4.

Missing Data

As shown in Table 3, 92.49% (n=7715) of the 8341 registered participants completed stress assessments prior to the OPC, whereas 30.09% (n=2510) completed the postassessment. Approximately one-quarter of the total sample (28.86%, n=2407) provided assessments both before and after their participation in the OPC. Most of these completers engaged in the platform at least once (89.78%, n=2161). Internal reliability on item-level data of the PSS-4 was high (Cronbach α=.82). Table 3 also presents perceived stress by completers and noncompleters.

Table 3. Perceived stress by completer status and time points.

Stress assessment	Did not complete both assessments (n=5934)	Completed both the pre and post PSS-4[a] (n=2407)	Total (N=8341)	*P* value
Preprogram stress level				<.001
N	5308	2407	7715	
Mean (SD)	6.0 (3.1)	5.6 (3.0)	5.9 (3.1)	
Median (range)	6.0 (0.0 to 16.0)	5.0 (0.0 to 16.0)	6.0 (0.0 to 16.0)	
Postprogram stress level				.002
N	103	2407	2510	
Mean (SD)	5.1 (2.5)	4.3 (2.7)	4.3 (2.7)	
Median (range)	5.0 (0.0 to 11.0)	4.0 (0.0 to 16.0)	4.0 (0.0 to 16.0)	
Stress reduction (difference from post to pre; negative value indicates stress reduction)				NA[b]
N	0	2407	2407	
Mean (SD)	NA	–1.3 (2.5)	–1.3 (2.5)	
Median (range)	NA	–1.0 (–11.0 to –9.0)	–1.0 (–11.0 to –9.0)	

[a]PSS-4: Perceived Stress Scale-4.

[b]NA: not applicable.

Engagement in Live and Recorded Training

Figure 1 shows that approximately 15% of completers engaged every day and approximately 10% did not engage at all. Above approximately 3 days of engagement, the rates of engagement leveled off with some minor spikes at around 15 and 20 days of engagement. Figure 2 shows a similar picture in the engagement pattern for those using only live programs.

Figure 1. Days of mindfulness practice among completers.

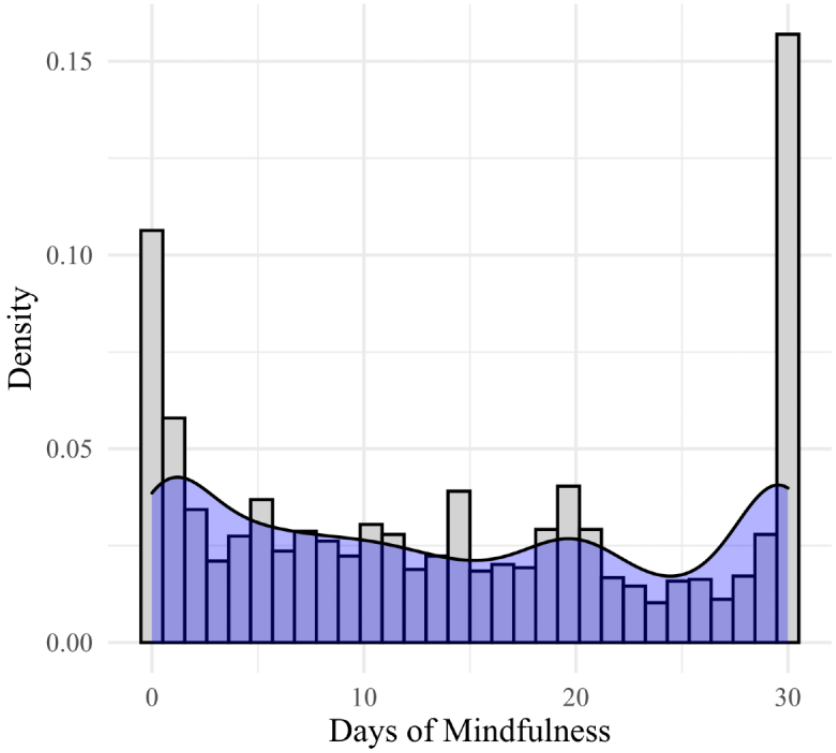
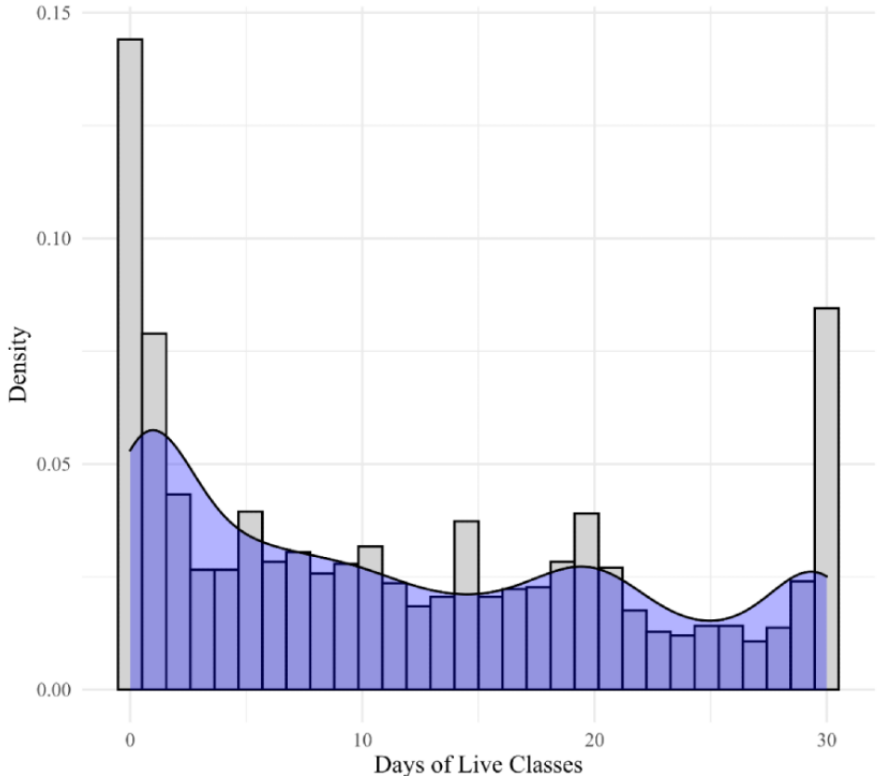
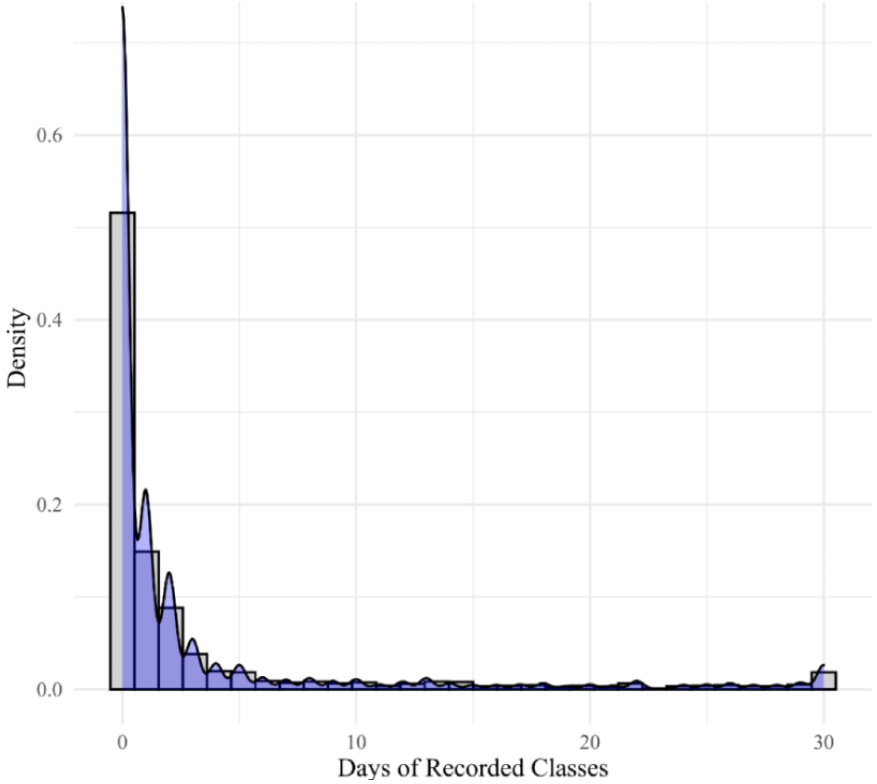


Figure 2. Completer engagement in live programs.



Use of recorded sessions presents a different story. As shown in Figure 3, many participants did not engage with recorded sessions at all (0 days: n=5652/8341, 67.76%), and for those who did, there was a steep decline of usage across the days. Most of those who used recordings at all (2689/8341, 32.22%) tended to use them 3-4 times at most. The remaining number of days using recordings was evenly scattered with a spike at 30 days (n=53 users).

Figure 3. Completer engagement in recorded programs.



Sampling Bias

We constructed a linear mixed-effects model predicting the PSS-4 score and examining completer status at baseline, allowing for random intercepts of PSS-4 items within person. Those who completed both pre- and post-OPC assessments showed a tendency to have lower stress scores at baseline measurement (B=–0.89, SE=0.02, $t_{10,134.13}$=–5.09; *P*<.001). This was a small but significant effect (Cohen *d*=–0.10). Given this potential source of bias, we included completer status as a control variable in all following analyses.

Change in Stress Levels

We next sought to examine the degree of change in stress from pre- to post-OPC. We built upon the previous model's random-effects structure by allowing for random slopes by time of measurement (pre- or post-OPC). We added time as a fixed effect and included completer status as a control variable in the model. We found a medium pre-post difference in stress levels (B=–0.32, SE=0.01, $t_{2393.25}$=–24.99; *P*<.001, *d*=–1.02). The estimates are summarized in Table 4 and Figure 4, and detailed model statistics are shown in Table 5.

Table 4. Fixed effects and their interactions on stress in the mixed-effects linear models.

Predictors	Model 1		Model 2		Model 3	
	Estimates, B (95% CI)	*P* value	Estimates, B (95% CI)	*P* value	Estimates, B (95% CI)	*P* value
Intercept	1.50 (1.48 to 1.52)	<.001	1.53 (1.50 to 1.55)	<.001	1.53 (1.50 to 1.55)	<.001
Time [Post]	–0.32 (–0.35 to –0.30)	<.001	–0.25 (–0.31 to –0.20)	<.001	–0.27 (–0.32 to –0.22)	<.001
Completer status	–0.10 (–0.13 to –0.06)	<.001	–0.06 (–0.10 to –0.02)	.002	–0.06 (–0.09 to –0.02)	.003
Days of mindfulness (DoM)	__[a]	—	–0.03 (–0.04 to –0.01)	<.001	—	—
Time×DoM	—	—	–0.03 (–0.05 to –0.01)	.007	—	—
Log of recorded days	—	—	—	—	–0.03 (–0.05 to –0.01)	.001
Log of live teaching days	—	—	—	—	–0.02 (–0.04 to –0.01)	.007
Time×log of recorded days	—	—	—	—	0.02 (0.00 to –0.04)	.09
Time×log of live teaching days	—	—	—	—	–0.03 (–0.05 to –0.01)	.002

[a]Not included in model.

Figure 4. Linear mixed effects model of One Percent Challenge (OPC) usage controlling for completer bias.

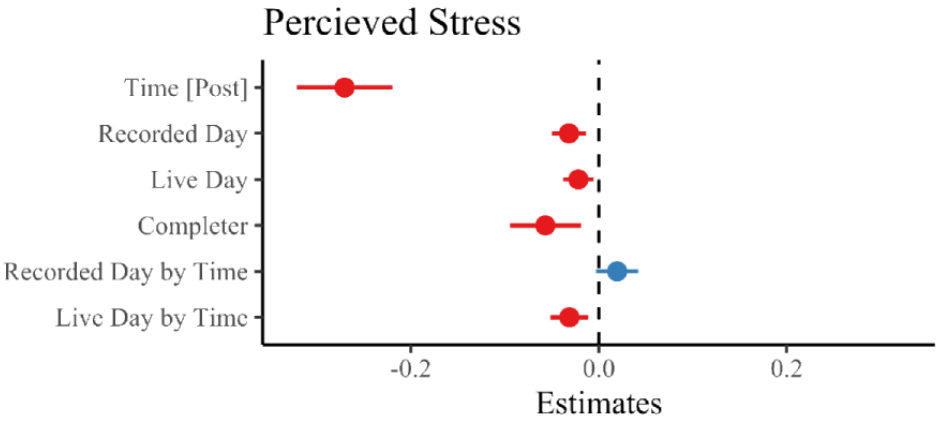


Table 5. Random effects on stress in the mixed-effects linear models.

Model statistic	Model 1	Model 2	Model 3
Variance (σ^2)	0.43	0.43	0.43
τ_{00}			
Question: (Time:ID)	0.07	0.07	0.07
Time:ID	0.07	0.07	0.07
ID	0.36	0.36	0.36
Intracorrelation coefficient	0.54	0.54	0.54
N			
Question	4	4	4
Time	2_{Time}	2_{Time}	2
ID	7281	7281	7281
Observations	40,984	40,984	40,984
Marginal R^2/conditional R^2	0.030/0.550	0.033/0.551	0.033/0.551

Impact of Days of Mindfulness on Stress Reduction

Days of mindfulness was highly positively skewed (skew=1.50; D'Augustino test for skewness, z=41.69; P<.001). Thus, we proceeded with a log-transformed version of the days of mindfulness variable and its two contributors, days of live sessions and days of recorded sessions, for all following analyses. We added days of mindfulness and the interaction of time point and days of mindfulness to the model to further explore its potential moderating impact. This model explained significantly more variance over and above the time point alone (χ^2_2=27.98, P<.001). In addition, this log-transformed model explained the data better than the model that left days of mindfulness untransformed (χ^2_0=7.59, P<.001).

Days of mindfulness had an impact on the reduction of stress over and above the general effects of engaging in the OPC (B=–0.03, SE=0.01, $t_{3277.22}$=–2.69; P=.007, d=–0.09). This represents an approximate 2.9% decrease in stress per day of mindfulness practice. Simply engaging in the OPC in some form over the time period of the OPC reliably reduced stress, and engaging over more days resulted in greater improvement. There was also some evidence that, accounting for all of these effects, individuals who were relatively less stressed overall were more likely to engage in more days of mindfulness (B=–0.03, SE=0.01, $t_{12,278.50}$=–3.49; P<.001, d=–0.06; see Table 4).

Relative Impact of Live Versus Recorded Programs on Stress Reduction

Breaking down days of mindfulness into days of live programs and days of recorded programs increased the explanatory power of the model (χ^2_2=9.06, P=.01). Participation in live programs drove decreases in stress (B=–0.03, SE=0.01, $t_{3258.61}$=–3.03; P=.002, d=–0.11), whereas participation in recorded classes did not (B=–0.02, SE=0.01, $t_{3908.17}$=1.79; P=.07, d=0.06; see Table 4). Participating in a live class resulted in a 3.1% reduction in stress.

Impact of Participant Demographics

To examine the additional impact of participant demographics, we removed those who reported that they would "Prefer not to respond" in response to gender (n=120) or race (n=544). We then removed a cluster of participants whose age was listed as "99" (n=24). This was notably above the next maximum reported age of 79; therefore, these responses were deemed spurious. Given the proportions in the sample, race was coded as anyone selecting "white" alone and those who selected anything else, including other options among "white." We pretested each demographic variable for differences in completer status to determine if there were additional sampling biases (see Table 6). Of these, women were marginally more likely to complete both time points of measurement (P=.09), and therefore we included the interaction of gender and completer status in the model to be conservative.

As shown in Table 6, we added the demographic variables to our previous model to evaluate their impact at baseline and their influence on change in stress over and above completer status and days of participating in live sessions or recorded sessions. Older participants had less stress at baseline (0.9% decrease in stress per year of age), with no clear difference in change in stress by age. Nonwhite participants had more stress at baseline (approximately 20.6% higher), but experienced greater reductions in stress from pre- to post-OPC (approximate 15.3% greater reduction) such that they had similar stress levels to those of the white participants post-OPC (who themselves experienced gains). The model statistics are summarized in Table 7.

Table 6. Demographic predictors of change in stress.

Predictors	Estimates, B (SE)	95% CI	*P* value (*df*=6893)
Intercept	1.91 (0.13)	1.66 to 2.17	<.001
Time [post]	–0.33 (0.11)	–0.56 to –0.11	.003
Completer status	–0.02 (0.05)	–0.12 to 0.07	.60
Gender [male]	0.18 (0.12)	–0.05 to 0.42	.12
Log of recorded days	–0.05 (0.02)	–0.08 to –0.02	.002
Log of live teaching days	0.00 (0.02)	–0.03 to 0.04	.86
Age	–0.01 (0.00)	–0.01 to 0.00	<.001
Race: Nonwhite	0.19 (0.06)	0.07 to 0.30	.001
Completer×Gender [male]	–0.25 (0.12)	–0.48 to –0.02	.03
Time×Log of recorded days	0.00 (0.02)	–0.03 to 0.03	.99
Time×Log of live teaching days	–0.06 (0.02)	–0.09 to –0.02	.006
Time×Gender [male]	–0.01 (0.05)	–0.11 to 0.10	.87
Time×Age	0.00 (0.00)	0.00 to 0.01	.06
Time×Nonwhite	–0.14 (0.05)	–0.24 to –0.04	.006

Table 7. Random effects of the model including demographic factors to predict changes in stress levels.

Model statistic	Value
σ^2	0.42
τ_{00}	
Question:(Time:ID)	0.08
Time:ID	0.05
ID	0.27
Intracorrelation coefficient	0.48
N	
Question	4
Time	2
ID	703
Observations	6911
Marginal R^2/conditional R^2	0.050/0.509

Discussion

Principal Findings

This study is the first to explore the relative contribution of live and recorded online mindfulness training to the reduction of stress. In this study of participants in the eMindful OPC, completers demonstrated a reduction in stress across their 30 days of participation with a medium effect size per Cohen *d* (95% CI 0.54-0.63). This is consistent with multiple meta-analyses of RCTs examining the impact of online mindfulness programs on perceived stress [3,29]. Perhaps more importantly, this is the first paper to report that participants' use of live teaching clearly drives the effects on stress reduction.

First, participating *at all* in eMindful OPC services is beneficial from a stress reduction perspective. Moreover, those who used live teaching and supplemented with recordings demonstrated the greatest effects, with an average improvement of 1.23 points on the PSS-4. Similarly, participating in only live teaching during the OPC conferred a smaller, but significant, reduction in stress. Although using only recorded programs or "on demand" content for practice did not reduce stress, the use of recordings did serve an important function. On average, participants used recordings only a handful of times across the month, suggesting that this minor usage is sufficient to supplement a regular daily mindfulness practice. Notably, running this same model using pure attendance numbers rather than days of practice did not show effects, indicating that changes in stress are more related to regular daily practice than the number of practices completed.

Second, in addition to benefits occurring from participating *at all* in the OPC, our data confirm that the more days one practices, the better. Stress reduction occurred in our sample from participating at all, and the remaining distance toward enduring change occurred at a rate of approximately 2.9% reduction per day of practice.

Why Might Live Online Teaching Drive Stress Reduction?

It makes intuitive sense that live programs would drive stress reduction, as the social component of live teaching, group discussion, and the importance of processing personal experiences are all critical elements in adult learning, particularly for higher-order thinking skills [2]. To sort through the high volume, rapidly changing, and often contradictory information coming at employees on a daily basis, higher-order thinking skills are imperative. In fact, such skills are thought to better prepare individuals to manage change than any knowledge or skill set: "The need to provide a highly educated, skilled workforce capable of providing solutions to 21st century challenges and issues has never been greater" [36]. The education literature can provide a better understanding of this phenomenon.

In academia, studies of online learning have mostly focused on learning through asynchronous platforms where participants access resources such as recorded lectures, readings, quizzes, and discussion boards at their convenience [37]. With the rapidly evolving technology of webcams and microphones to allow real-time participation, synchronous online learning has been increasing. Although there are still few studies on the topic, the research performed to date shows that participants indicate that any possible frustration with the technology itself is overridden by the convenience of synchronous online learning [37]. This literature will likely evolve rapidly given the move from offline to online education in a short period during the COVID-19 pandemic [38], which will hopefully provide even greater insight into online learning.

The peer-reviewed literature on the effectiveness of online learning to date can help explain why the live teaching component is the driver of our outcomes. Researchers of online learning have long supported the value of social interaction as a crucial element in the learning process [39]. Despite some equivocal evidence, successful synchronous courses provide opportunities for participants to share experiences and interact with others. This opportunity is not available in recorded programs or on static apps. Live teaching allows for peer-to-peer as well as peer-to-instructor interaction, a crucial element in the learning process. Research shows that peer-to-instructor interaction remains strong in online synchronous learning, sometimes superior to in-person learning, and that there is greater variability regarding peer-to-peer interaction [37]. The quality of peer-to-peer interaction depends in part on the program's use of polling and chats in addition to microphones that allow participants to express their thoughts and experiences. Learning theory clarifies that adults use personal life experiences as a framework for all subsequent learning [40]. Processing their mindfulness experiences together in class provides opportunities for participants to connect to their current learning. This cognitive scaffolding is crucial for higher-order thinking and occurs through interaction.

Only synchronous online programs allow for discussion with others in the class as well as with the instructor. Since participants are processing their own experiences in real time, each comment can shift the direction of thinking in another person. Owing to this constant shifting, individuals tend to reach higher levels of thinking as they respond to each other's statements [39]. Discussion and group input thus invite individuals to reach higher levels of thinking, which may be important in navigating the constant change and uncertainty required in today's world.

The education literature can provide more information regarding adult learners. There is also a wide range of participant confidence in contributing online, with some feeling less confident to contribute online versus in person and others expressing more confidence to contribute online [37]. Whether use of the online synchronous classroom tends to engender more or less confidence in participating may depend on the level of interactive opportunities the program provides. When there is reduced interaction online versus in person, a likely explanation has to do with how technology is used within sessions. The technology has developed to allow for considerable peer-to-peer interaction, with features such as polls, surveys, and breakout rooms; however, in practice, the interactive technology is not always used to its full potential. Ng and Jeffery [41] found that teachers using new software tended to stick to the traditional didactic lecture style of teaching, estimating that they devoted a quarter of the time to interactive activities online than they would in a comparable classroom. Fortunately, the competitive nature of the corporate world has demanded creative methods of engagement from onboarding through delivery of each session. OPC sessions employ a small didactic nugget with inquiry that invites participants to place the lesson in daily context. Most of the session time is focused on experiential learning, community practice, and processing the experience.

Observations from the industrial organizational literature also support the importance of a live teacher. In studies on performance feedback, researchers note the importance of learners perceiving "social presence," a sense of human warmth and being with another person [42]. In addition to influencing the perceived utility of feedback systems in learning, social presence appears to be particularly important in distance-based learning [42]. Not surprisingly, human feedback is perceived as having greater warmth and sense of closeness than machine-generated feedback. Similarly, the richer the feedback media (eg, containing visual and auditory input), the greater the perceived social presence as it contains greater social cues [42]. Similar observations have been demonstrated in the behavioral change literature. For example, both human-generated email feedback and computer-tailored feedback to participants in a weight loss trial appeared to have the same impact at 3 months, but only the improvement from the human-generated email group was maintained at 6 months [43,44]. The use of recorded practices and computer-tailored models is likely to be quite nuanced. It is clear that multiple variables impact behavior change and learning outcomes from online participants, such as the education of participants [44]. The education literature

also shows the potential impact of age on one's preference for live online versus recorded teaching. Depending on the context, student preferences for recorded training versus live online training have been noted, particularly for older students in situations of content learning [45]. Our results did not show age as a moderator of our finding that the live programs, supplemented by on-demand sessions, is what drove the reduction in stress.

Although eMindful's focus has been live teaching through a virtual classroom, the company developed an app and recorded options for daily use on the platform to further support practice. Our findings suggest that participants obtained the most benefit when using the on-demand options as a supplement to support live practice. This effect has been detected in at least one other trial as well. In an RCT comparing participants with access to daily guided meditations to participants who had the same access but also received a 1-hour web-based live training session for 6 weeks, Wahbeh and Oken [46] showed that usage of the live web platform increased daily meditation practice.

Limitations of the Study

This study has several limitations that should be considered in weighing the findings. First, the sample for this study came from a large pool of participants in a particular program (the OPC) at a single mindfulness training company. Hence, it is unclear if the findings would generalize to other programs in the company or to other companies. Second, there is a considerable amount of missing data in the sample. Analyzing data from corporate programs is inherently challenging, since each corporate client dictates the data to be collected, providing considerable nonrandom variability. For example, missingness for sociodemographic data is a result of when the participants enrolled, the processes at eMindful at the time, and each organization's preferences. Hence, the demographics provided cannot be assumed to be a random representation of the entire database, or of corporate employees in general. Similarly, missingness for outcomes data appears to be nonrandom, making it extremely important that the analyses allowed for potential sources of bias. Fortunately, our analyses did take potential sources of bias into consideration. Completers showed a 0.045-point lower pre-OPC stress level compared with that of noncompleters. This completion bias suggests that noncompleters were slightly more stressed. Based on these data, we would expect that if these noncompleters did complete both pre- and post-OPC measures, the effects observed on outcomes would only increase. However, this may depend on engaging them more in services. Per these analyses, if all people who engaged in services completed both surveys, we may reasonably assume that the models presented regarding the reduction of stress are conservative estimates of the potential of the OPC to reduce stress. Finally, the baseline rates of using recorded sessions were much lower than those of using live sessions. We do not know what impact this had on our results. We attempted to account for this bias by analyzing log-transformed variables and including the effect of engaging in recorded sessions at all. Although our findings may speak to ecologically valid situations, the field would benefit from additional study of well-controlled comparisons of recorded versus live sessions that compel equal dosage in a randomized fashion.

Conclusion

Participation *at all* with the eMindful OPC reduced stress, and live online teaching drove this outcome. Regular practice across the month led to an even greater reduction in stress. The use of recorded or on-demand offerings to supplement live practices conferred the strongest likelihood of achieving a clinically significant change in stress levels. Our findings are in stark contrast to the rapid evolution of online mindfulness training for the workplace. Specifically, the market is reproducing apps and recorded teaching at an unprecedented pace, whereas our results demonstrate that live teaching leads to greater stress reduction.

Acknowledgments

The authors wish to thank eMindful Inc for providing the anonymized data for this study, for funding time for RW and DS to work on the paper, and for covering publication fees. The authors also appreciate the support of Mary Pigatti, CEO of eMindful, Inc, and Zev Suissa, MBA, CIO at eMindful, Inc, for review of the manuscript and input on the industry. The authors appreciate Himanshu Chaudhary, CTO of eMindful Inc, for support with data acquisition, and Kelley McCabe, MBA for recognizing the literature gap.

Authors' Contributions

RW conceived of the study design with substantial contributions from MF. DS oversaw acquisition of data. MF performed all statistical analyses. MF and RW interpreted the data. RW drafted the article. All authors provided critical revision of the article, and reviewed and approved the final manuscript.

Conflicts of Interest

RW serves as Chief Science Officer for eMindful Inc, serves on the Clinical Advisory Board, owns stock and stock options of the company (1.7%), and has helped to developed curriculum for eMindful's training programs over the past decade, some of which were accessed by participants using the eMindful platform under study. eMindful Inc asked RW to design a study to evaluate the benefits of live mindfulness training versus use of recorded training options, and had no input into the specific design, conduct of the analysis, or interpretation of the data. eMindful provided no prohibitions or requirements regarding what could be

published. In the past 5 years, MF was paid for analyses by eMindful Inc; however, MF has no conflict to declare relevant to this study. DS works full time for eMindful Inc.

References

1. Simonsson O, Fisher S, Martin M. Awareness and experience of mindfulness in Britain. Sociol Res Online 2020 Dec 27;26(4):833-852. [doi: 10.1177/1360780420980761]
2. Simonsson O, Martin M, Fisher S. Sociodemographic characteristics and health status of mindfulness users in the United States. Mindfulness 2020 Aug 21;11(12):2725-2729. [doi: 10.1007/s12671-020-01486-4]
3. Spijkerman M, Pots W, Bohlmeijer E. Effectiveness of online mindfulness-based interventions in improving mental health: a review and meta-analysis of randomised controlled trials. Clin Psychol Rev 2016 Apr;45:102-114 [FREE Full text] [doi: 10.1016/j.cpr.2016.03.009] [Medline: 27111302]
4. The employer investment in employee well-being: Tenth Annual Employer-Sponsored Health and Well-being Survey. Business Group on Health. 2020. URL: https://www.businessgrouphealth.org/en/resources/the-employer-investment-in-employee-well-being-tenth-annual-employer-sponsored-health [accessed 2022-01-12]
5. Wolever RQ, Schwartz ER, Schoenberg PL. Mindfulness in corporate America: is the Trojan Horse ethical? J Altern Complement Med 2018 May;24(5):403-406. [doi: 10.1089/acm.2018.0171] [Medline: 29717881]
6. Asmundson G, Taylor S. Coronaphobia revisted: A state-of-the-art on pandemic-related fear, anxiety, and stress. J Anxiety Disord 2020 Dec;76:102326 [FREE Full text] [doi: 10.1016/j.janxdis.2020.102326] [Medline: 33142124]
7. Fofana NK, Latif F, Sarfraz S, Bilal, Bashir MF, Komal B. Fear and agony of the pandemic leading to stress and mental illness: An emerging crisis in the novel coronavirus (COVID-19) outbreak. Psychiatry Res 2020 Sep;291:113230 [FREE Full text] [doi: 10.1016/j.psychres.2020.113230] [Medline: 32593067]
8. Sinclair RR, Allen T, Barber L, Bergman M, Britt T, Butler A, et al. Occupational health science in the time of COVID-19: now more than ever. Occup Health Sci 2020 Jun 01;4(1-2):1-22 [FREE Full text] [doi: 10.1007/s41542-020-00064-3] [Medline: 32838031]
9. Trougakos JP, Chawla N, McCarthy JM. Working in a pandemic: exploring the impact of COVID-19 health anxiety on work, family, and health outcomes. J Appl Psychol 2020 Nov;105(11):1234-1245. [doi: 10.1037/apl0000739] [Medline: 32969707]
10. Carvalheiro J, Conceição VA, Mesquita A, Seara-Cardoso A. Acute stress impairs reward learning in men. Brain Cogn 2021 Feb;147:105657 [FREE Full text] [doi: 10.1016/j.bandc.2020.105657] [Medline: 33341656]
11. Schwabe L, Wolf OT. Learning under stress impairs memory formation. Neurobiol Learn Mem 2010 Feb;93(2):183-188. [doi: 10.1016/j.nlm.2009.09.009] [Medline: 19796703]
12. Sharma M, Rush SE. Mindfulness-based stress reduction as a stress management intervention for healthy individuals: a systematic review. J Evid Based Complementary Altern Med 2014 Oct 22;19(4):271-286. [doi: 10.1177/2156587214543143] [Medline: 25053754]
13. Fjorback L, Arendt M, Ornbøl E, Fink P, Walach H. Mindfulness-based stress reduction and mindfulness-based cognitive therapy: a systematic review of randomized controlled trials. Acta Psychiatr Scand 2011 Aug;124(2):102-119. [doi: 10.1111/j.1600-0447.2011.01704.x] [Medline: 21534932]
14. Pascoe MC, Thompson DR, Jenkins ZM, Ski CF. Mindfulness mediates the physiological markers of stress: Systematic review and meta-analysis. J Psychiatr Res 2017 Dec;95:156-178. [doi: 10.1016/j.jpsychires.2017.08.004] [Medline: 28863392]
15. Heckenberg RA, Eddy P, Kent S, Wright BJ. Do workplace-based mindfulness meditation programs improve physiological indices of stress? A systematic review and meta-analysis. J Psychosom Res 2018 Nov;114:62-71. [doi: 10.1016/j.jpsychores.2018.09.010] [Medline: 30314581]
16. Wolever RQ, Bobinet KJ, McCabe K, Mackenzie ER, Fekete E, Kusnick CA, et al. Effective and viable mind-body stress reduction in the workplace: a randomized controlled trial. J Occup Health Psychol 2012 Apr;17(2):246-258. [doi: 10.1037/a0027278] [Medline: 22352291]
17. Hofmann SG, Sawyer AT, Witt AA, Oh D. The effect of mindfulness-based therapy on anxiety and depression: a meta-analytic review. J Consult Clin Psychol 2010 Apr;78(2):169-183 [FREE Full text] [doi: 10.1037/a0018555] [Medline: 20350028]
18. Khoury B, Lecomte T, Fortin G, Masse M, Therien P, Bouchard V, et al. Mindfulness-based therapy: a comprehensive meta-analysis. Clin Psychol Rev 2013 Aug;33(6):763-771. [doi: 10.1016/j.cpr.2013.05.005] [Medline: 23796855]
19. Blanck P, Perleth S, Heidenreich T, Kröger P, Ditzen B, Bents H, et al. Effects of mindfulness exercises as stand-alone intervention on symptoms of anxiety and depression: systematic review and meta-analysis. Behav Res Ther 2018 Mar;102:25-35. [doi: 10.1016/j.brat.2017.12.002] [Medline: 29291584]
20. Good DJ, Lyddy CJ, Glomb TM, Bono JE, Brown KW, Duffy MK, et al. Contemplating mindfulness at work. J Manage 2015 Nov 19;42(1):114-142. [doi: 10.1177/0149206315617003]
21. Lomas T, Medina JC, Ivtzan I, Rupprecht S, Eiroa-Orosa FJ. Mindfulness-based interventions in the workplace: An inclusive systematic review and meta-analysis of their impact upon wellbeing. J Positive Psychol 2018 Sep 30;14(5):625-640. [doi: 10.1080/17439760.2018.1519588]

22. Bartlett L, Martin A, Neil AL, Memish K, Otahal P, Kilpatrick M, et al. A systematic review and meta-analysis of workplace mindfulness training randomized controlled trials. J Occup Health Psychol 2019 Feb;24(1):108-126. [doi: 10.1037/ocp0000146] [Medline: 30714811]
23. Wolever RQ, Kahn JA, Davis J, Shields D, Schoenberg PLA. Introducing the eMCC™: a validated taxonomy to advance targeted application of mindfulness skills. Mindfulness 2019 Dec 19;11(3):698-708. [doi: 10.1007/s12671-019-01280-x]
24. Sumantry D, Stewart KE. Meditation, mindfulness, and attention: a meta-analysis. Mindfulness 2021 Feb 02;12(6):1332-1349. [doi: 10.1007/s12671-021-01593-w]
25. Yakobi O, Smilek D, Danckert J. The effects of mindfulness meditation on attention, executive control and working memory in healthy adults: a meta-analysis of randomized controlled trials. Cogn Ther Res 2021 Jan 04;45(4):543-560. [doi: 10.1007/s10608-020-10177-2]
26. Jamieson SD, Tuckey MR. Mindfulness interventions in the workplace: a critique of the current state of the literature. J Occup Health Psychol 2017 Apr;22(2):180-193. [doi: 10.1037/ocp0000048] [Medline: 27643606]
27. Vonderlin R, Biermann M, Bohus M, Lyssenko L. Mindfulness-based programs in the workplace: a meta-analysis of randomized controlled trials. Mindfulness 2020 Mar 02;11(7):1579-1598. [doi: 10.1007/s12671-020-01328-3]
28. Mesmer-Magnus J, Manapragada A, Viswesvaran C, Allen JW. Trait mindfulness at work: a meta-analysis of the personal and professional correlates of trait mindfulness. Hum Perform 2017 Apr 21;30(2-3):79-98. [doi: 10.1080/08959285.2017.1307842]
29. Jayewardene WP, Lohrmann DK, Erbe RG, Torabi MR. Effects of preventive online mindfulness interventions on stress and mindfulness: a meta-analysis of randomized controlled trials. Prev Med Rep 2017 Mar;5:150-159 [FREE Full text] [doi: 10.1016/j.pmedr.2016.11.013] [Medline: 28050336]
30. Mrazek AJ, Mrazek MD, Cherolini CM, Cloughesy JN, Cynman DJ, Gougis LJ, et al. The future of mindfulness training is digital, and the future is now. Curr Opin Psychol 2019 Aug;28:81-86. [doi: 10.1016/j.copsyc.2018.11.012] [Medline: 30529975]
31. Singh P. Unhooking the drama: meditation app statistics to know in 2021. appinventiv. 2020 Feb 28. URL: https://appinventiv.com/blog/latest-meditation-app-statistics/ [accessed 2021-07-02]
32. Schultchen D, Terhorst Y, Holderied T, Stach M, Messner E, Baumeister H, et al. Stay present with your phone: a systematic review and standardized rating of mindfulness apps in European app stores. Int J Behav Med 2021 Oct 20;28(5):552-560 [FREE Full text] [doi: 10.1007/s12529-020-09944-y] [Medline: 33215348]
33. Mani M, Kavanagh DJ, Hides L, Stoyanov SR. Review and evaluation of mindfulness-based iPhone apps. JMIR Mhealth Uhealth 2015 Aug 19;3(3):e82 [FREE Full text] [doi: 10.2196/mhealth.4328] [Medline: 26290327]
34. Gál É, tefan S, Cristea IA. The efficacy of mindfulness meditation apps in enhancing users' well-being and mental health related outcomes: a meta-analysis of randomized controlled trials. J Affect Disord 2021 Jan 15;279:131-142. [doi: 10.1016/j.jad.2020.09.134] [Medline: 33049431]
35. Cohen S, Kamarck T, Mermelstein R. A global measure of perceived stress. J Health Soc Behav 1983 Dec;24(4):385. [doi: 10.2307/2136404]
36. Doerfert D. National Research Agenda: American Association for Agricultural Education's research priority areas for 2011-2015. Lubbock, TX: Texas Tech University, Department of Agricultural Education and Communications; 2011. URL: http://www.aaaeonline.org/resources/Documents/AAAE%20National%20Research%20Agenda.pdf [accessed 2022-01-12]
37. Yates J. Synchronous online CPD: empirical support for the value of webinars in career settings. Br J Guid Couns 2014 Feb 06;42(3):245-260. [doi: 10.1080/03069885.2014.880829]
38. Liu X, Zhou J, Chen L, Yang Y, Tan J. Impact of COVID-19 epidemic on live online dental continuing education. Eur J Dent Educ 2020 Nov 30;24(4):786-789 [FREE Full text] [doi: 10.1111/eje.12569] [Medline: 32648989]
39. Brierton S, Wilson E, Kistler M, Flowers J, Jones D. A comparison of higher order thinking skills demonstrated in synchronous and asynchronous online college discussion posts. Nacta J 2016;60(1):14 [FREE Full text]
40. Knowles M, Holton IE, Swanson R. The Adult Learner: The definitive classic in adult education and human resource development. 6th edition. New York: Elsevier; 2005.
41. Ng DM, Jeffery RW. Relationships between perceived stress and health behaviors in a sample of working adults. Health Psychol 2003 Nov;22(6):638-642. [doi: 10.1037/0278-6133.22.6.638] [Medline: 14640862]
42. Walter N, Ortbach K, Niehaves B. Designing electronic feedback – Analyzing the effects of social presence on perceived feedback usefulness. Int J Hum Comput Stud 2015 Apr;76:1-11. [doi: 10.1016/j.ijhcs.2014.12.001]
43. Tate DF, Jackvony EH, Wing RR. A randomized trial comparing human e-mail counseling, computer-automated tailored counseling, and no counseling in an Internet weight loss program. Arch Intern Med 2006 Aug 14;166(15):1620-1625. [doi: 10.1001/archinte.166.15.1620] [Medline: 16908795]
44. de Vries H, Logister M, Krekels G, Klaasse F, Servranckx V, van Osch L. Internet based computer tailored feedback on sunscreen use. J Med Internet Res 2012 Apr 30;14(2):e48 [FREE Full text] [doi: 10.2196/jmir.1902] [Medline: 22547528]
45. Islam M, Kim D, Kwon M. A comparison of two forms of instruction: pre-recorded video lectures vs. live ZOOM lectures for education in the business management field. Sustainability 2020 Oct 02;12(19):8149. [doi: 10.3390/su12198149]
46. Wahbeh H, Oken BS. Internet mindfulness meditation intervention for the general public: pilot randomized controlled trial. JMIR Ment Health 2016 Aug 08;3(3):e37 [FREE Full text] [doi: 10.2196/mental.5900] [Medline: 27502759]

Abbreviations

OPC: One Percent Challenge
PSS-4: Perceived Stress Scale-4
RCT: randomized controlled trial

Edited by A Mavragani; submitted 09.07.21; peer-reviewed by M Nahum; comments to author 01.08.21; revised version received 21.09.21; accepted 22.12.21; published 21.01.22.

Please cite as:
Wolever RQ, Finn MTM, Shields D
The Relative Contributions of Live and Recorded Online Mindfulness Training Programs to Lower Stress in the Workplace: Longitudinal Observational Study
J Med Internet Res 2022;24(1):e31935
URL: https://www.jmir.org/2022/1/e31935
doi:10.2196/31935
PMID:35060911

Original Paper

Using Intervention Mapping to Develop a Decision Support System–Based Smartphone App (selfBACK) to Support Self-management of Nonspecific Low Back Pain: Development and Usability Study

Malene Jagd Svendsen[1,2], MSc; Louise Fleng Sandal[2], PhD; Per Kjær[2,3], PhD; Barbara I Nicholl[4], PhD; Kay Cooper[5], PhD; Frances Mair[4], MD, PhD; Jan Hartvigsen[2,6], PhD; Mette Jensen Stochkendahl[2,6], PhD; Karen Søgaard[2,7], PhD; Paul Jarle Mork[8], PhD; Charlotte Rasmussen[1], PhD

[1]The National Research Centre for the Working Environment, Copenhagen, Denmark

[2]Department of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark

[3]Health Sciences Research Centre, UCL University College, Odense, Denmark

[4]Institute of Health and Wellbeing, General Practice & Primary Care, University of Glasgow, Glasgow, United Kingdom

[5]School of Health Sciences, Robert Gordon University, Aberdeen, United Kingdom

[6]Chiropractic Knowledge Hub, Odense, Denmark

[7]Department of Clinical Research, University of Southern Denmark, Odense, Denmark

[8]Department of Public Health and Nursing, Norwegian University of Science and Technology, Trondheim, Norway

Corresponding Author:
Malene Jagd Svendsen, MSc
The National Research Centre for the Working Environment
Lersø Parkallé 105
Copenhagen, 2100
Denmark
Phone: 45 20259734
Email: mas@nfa.dk

Abstract

Background: International guidelines consistently endorse the promotion of self-management for people with low back pain (LBP); however, implementation of these guidelines remains a challenge. Digital health interventions, such as those that can be provided by smartphone apps, have been proposed as a promising mode of supporting self-management in people with chronic conditions, including LBP. However, the evidence base for digital health interventions to support self-management of LBP is weak, and detailed descriptions and documentation of the interventions are lacking. Structured intervention mapping (IM) constitutes a 6-step process that can be used to guide the development of complex interventions.

Objective: The aim of this paper is to describe the IM process for designing and creating an app-based intervention designed to support self-management of nonspecific LBP to reduce pain-related disability.

Methods: The first 5 steps of the IM process were systematically applied. The core processes included literature reviews, brainstorming and group discussions, and the inclusion of stakeholders and representatives from the target population. Over a period of >2 years, the intervention content and the technical features of delivery were created, tested, and revised through user tests, feasibility studies, and a pilot study.

Results: A behavioral outcome was identified as a proxy for reaching the overall program goal, that is, increased use of evidence-based self-management strategies. Physical exercises, education, and physical activity were the main components of the self-management intervention and were designed and produced to be delivered via a smartphone app. All intervention content was theoretically underpinned by the behavior change theory and the normalization process theory.

Conclusions: We describe a detailed example of the application of the IM approach for the development of a theory-driven, complex, and digital intervention designed to support self-management of LBP. This description provides transparency in the developmental process of the intervention and can be a possible blueprint for designing and creating future digital health interventions for self-management.

(J Med Internet Res 2022;24(1):e26555) doi:10.2196/26555

KEYWORDS

intervention mapping; behavior change; low back pain; self-management; mHealth; app-based intervention; decision support system; digital health intervention; mobile phone

Introduction

Low back pain (LBP) is well-documented as one of the most common reasons for activity limitation, sick leave, and disability [1-4]. Clinical guidelines for LBP consistently endorse patient education, general physical activity, exercise, and the promotion of core self-management components of frontline care [5-9]. However, implementation of self-management may be challenging, perhaps because of its multifaceted complex nature, with several interacting components and health care settings. Therefore, new effective ways of delivering supported self-management for people with LBP are needed.

Digital health interventions (DHIs), such as those that can be provided by smartphone apps, have been proposed as a promising mode for supporting self-management in people with chronic conditions [10]. In a recent systematic review on the use of DHIs for supporting self-management of LBP, we found that the literature was heterogeneous in terms of reporting intervention details, making it difficult to understand what might work best, for whom, and in what circumstances [11]. The descriptions of the intervention development and use of theory were either brief or completely lacking in all the included studies, and the evidence base for DHIs to support self-management of LBP was weak [11]. In another systematic review, Garg et al [12] found that web-based interventions for supporting individuals with LBP were useful. In particular, interventions that offer feedback or tailoring based on user responses and elements from cognitive behavioral therapy seem to be beneficial; however, the effectiveness of DHIs in supporting self-management of LBP remains unclear [12]. A systematic review of smartphone apps for self-management of LBP concluded that researchers, health care professionals (HCPs), people with LBP, and app developers need to work closely together to develop DHIs that are accurate, evidence-based, and engaging [13]. Structured intervention mapping (IM) provides a well-defined framework for the development, implementation, and evaluation of interventions by integrating the target population and stakeholders in the process [14]. Using IM for development also fits with the Medical Research Council framework for evaluation of complex interventions [15] and CONSORT-EHEALTH (Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth) [16] guidelines for reporting DHIs. One of the issues emphasized by CONSORT-EHEALTH is that a detailed description and documentation of the intervention is required to fully understand its effectiveness [16]. The stepwise process of IM provides just that, and it has previously been used to develop and adapt evidence-based self-management programs in other settings [17,18]. selfBACK [19], a case-based reasoning (CBR) DHI, which is delivered as an app, is designed to improve the self-management of nonspecific LBP to reduce pain-related disability. The selfBACK app provides weekly tailored self-management plans (SMPs) targeting physical activity, strength and flexibility exercises, and education. In addition, the app also provides access to a variety of tools and information on the management of LBP that the participants can use at their convenience. A randomized controlled trial (RCT) is examining the effectiveness of the intervention as an add-on to usual care [20]. The aim of this paper is to describe the IM process used to design and create selfBACK.

Methods

IM Approach

IM is a 6-step process for the development, implementation, and evaluation of an intervention; however, here we have described just 5 steps as step 6, the planning of the evaluation of the intervention, is published as a protocol for an RCT [20] alongside an implementation and process evaluation protocol [21]. Each step comprises several tasks, which, once completed, inform the next step, as detailed by Bartholomew et al [14]. The IM approach considers the target population, its surrounding environment, and the persons in that environment who might influence the target population. It aims to facilitate participation and consultation of stakeholders and provides a structure for the integration of theory, findings from empirical literature, and information collected from the target population. Throughout all the steps of the development phase, we used *core processes* (ie, posing questions, brainstorming, reviewing empirical literature, reviewing theories, assessing needs for new data, and developing a working list of answers) as the underlying methods [14].

Step 1: Logic Model of the Problem

First, we established a planning group comprising partners from both clinical and research backgrounds and those with expertise in app development and health innovation management. The planning group comprised physiotherapists, chiropractors, physicians, exercise physiologists, behavioral scientists, computer scientists, and app designers.

We conducted a needs assessment from a societal and user perspective as a response to a European Union Horizon 2020 program calling for “self-management of health and disease and decision support systems based on predictive computer modelling used by the person him or herself.” To meet the criteria in the call, the project aimed solely for an intervention targeting individuals with LBP and not their surrounding environment or the persons in that environment. The needs assessment was informed by 2 systematic literature reviews [11,22] that we conducted on digital interventions for supporting self-management of LBP, reviews of clinical practice guidelines, extensive supplementary literature searches on factors associated with LBP and self-management, interviews with patients with

LBP and HCPs, clinical experience, and group discussions in the planning group.

The process resulted in a logic model of the problem that mapped personal determining factors of LBP self-management, adverse behavior in relation to self-management and how that behavior affects LBP and associated health issues, and, finally, the impact on quality of life. Subsequently, based on recommendations from the LBP literature and previous experiences within the planning group with interventions targeting LBP, the overall program goal of the intervention was formulated.

Step 2: Program Outcomes and Objectives—Logic Model of Change

The program outcomes were formulated based on the scrutiny of systematic reviews, overview papers, expert opinion papers on core outcome domains, and outcome measures used in previous LBP interventions. To identify behavior-related outcomes, we consulted the literature on behavior change theory related to pain, pain-related disability, and self-management. We also used information from the 2 systematic reviews that we conducted on digital interventions to support self-management of LBP [11,22]. Building on the results from step 1 and an examination of the literature, a logic model of change specifying the outcomes and objectives was developed. The model included a definition of the change in behavior needed for improvement in LBP (*behavioral outcomes*) and detailed specifications of what people with LBP need to do to perform that behavior (*performance objectives*). Finally, a matrix pairing the performance objectives with the determining factors of LBP self-management identified in step 1, with positive statements about what needs to occur to achieve the performance objectives (*change objectives* [14]), was created.

Step 3: Program Design

In step 3, the intervention concepts were outlined in an iterative process of matching relevant behavioral change theory to the practical application ideas. Clinical practice guidelines and their accompanying patient leaflets, as well as pain management websites deemed to originate from trustworthy sources, were reviewed to outline the intervention content. The choice of specific physical exercises and educational content was determined in a parallel process with the GLA:D Back project, a group education and exercise program that translates guideline recommendations into a clinician-delivered package for the promotion of self-management in people with persistent or recurrent back pain [23]. A psychologist provided feedback on the educational content. The process of identifying theories of behavior change that would guide the choice of methods and practical applications of these methods was informed by the review of behavior change theory from step 2 and the extensive experience with trials on self-management of LBP among the planning group. Parts of the interviews conducted for step 1 provided valuable information on users' needs and wishes regarding the self-management app. We also purposefully reviewed the literature on practical ideas for enhancing engagement with the app, for example, as gamification and notification systems. For all determinants and performance objectives, theoretically underpinned behavior change techniques were chosen, and the practical application of each technique was brainstormed, discussed, and refined by the planning group. The outcome of this third step was an outline of the intervention themes, components, and sequences of the intervention.

Step 4: Program Production

All technical and practical features of delivering the selfBACK intervention through a smartphone app and the participant documents for use in the RCT study were created in step 4 [24,25]. Throughout the production process, brainstorming sessions and workshops among planning group members for production of content were held continuously, and technical solutions for running the digital intervention (ie, tailored decision support for self-management using artificial intelligence) were developed, tested, and refined. Furthermore, the user interface was designed, starting with wireframes, visual identity and design, and functional requirements. Before we designed the app's visual identity, we performed a search of existing apps on Google Play and the Apple App Store. The apps were qualitatively evaluated using Apple Human Interface Guidelines, Apple Research Kit, and Google's Material Design. These guidelines are considered state-of-the-art documents and trendsetters in app design. As part of the design process, we created fictional characters (personas) based on the literature scrutiny from the previous steps and the interviews conducted in step 2. The personas helped the app developers understand the needs, experiences, behaviors, and goals of patients with LBP and guide the design process. The user journey map was matched with the personas. This map served as a blueprint for the design and development phases of the app. The map depicted an overview of the journey that a user would embark on from experiencing LBP for the first time to finding an individual solution to their pain and how that journey may be guided toward using the selfBACK app.

A total of 2 feasibility studies with target population participants were conducted: 1 in the United Kingdom and 1 in Norway. The UK feasibility study (n=16) was conducted to explore the feasibility and acceptability of the baseline questionnaire, physical activity monitoring, and feedback strategies with a prototype app, whereas the Norwegian feasibility study (n=10) tested the full intervention with an early version of the selfBACK app [26]. In both the UK and Norwegian feasibility studies, we applied quantitative and qualitative methods to inform further development of the intervention. Simultaneously, app prototypes were tested by the planning group members and external users conducted both as group sessions and as real-time tests of app use in consecutive periods (n=65 female, 60 male). Finally, a pilot study in Denmark and Norway (n=51) was conducted using a complete version of the app to test recruitment and screening procedures and inform the design of the effect and process evaluation [27]. Intervention content was developed and continually refined in an iterative process over a period of >2 years. To ensure consistency throughout the app content and the trial documents, everything was first completed in English and then translated to Danish and Norwegian (settings for the RCT study).

Step 5: Implementation Plan

The plan for the implementation and adoption of the selfBACK app was based on the results from previous steps, the pilot study, and theory, taking into account both behavior change and implementation theories, to improve the likelihood of embedding the DHI into daily routines. The program use outcomes (adoption and implementation) were specified. The selfBACK app targets care-seeking patients; therefore, it was necessary to also consider the adoption of the selfBACK app from the recruiting HCPs' perspective. We used normalization process theory (NPT) to identify determinants of adoption and implementation and linked them to each performance objective. Finally, a matrix was created by planning the group discussions. We used behavior change techniques (BCTs) to convert the change objectives into practical strategies. Recruitment material for patients with LBP (app users) and recruiting clinicians was produced together with a plan of how to attract both groups to the project. Procedures for the recruitment of and initial and sustained contact with the HCPs were established, as well as the procedures for inclusion, screening, randomization, follow-up, and evaluation of the app users [20].

Results

Step 1: Logic Model of the Problem

In step 1, we interviewed 8 patients with LBP (5 [63%] men and 3 [27%] women) about their experience with treatment of LBP and how they usually self-managed their LBP. From these interviews, we identified and prioritized *not following evidence-based self-management strategies* as the most important changeable, adverse behavioral factor contributing to poor outcomes for people with LBP. This behavior is believed to be affected by the following personal determining factors: not being aware of ways of self-managing LBP (*lack of knowledge and awareness*); not possessing skills or being insecure about the ability to self-manage LBP (*low self-efficacy*); or fearing an increase in pain when doing physical activities (*fear-avoidance behavior and catastrophizing*), negative expectations about the course of LBP (*low outcome expectations*), and challenges encountered when trying to fit self-management strategies within the context of daily life (*low motivation*). In addition, the 2 interviewed HCPs reported that many patients relied on HCPs to cure their pain and that getting people to change their perception of how best to manage LBP (eg, avoid bed rest and stay active) was one of the biggest problems for HCPs (patients' *lack of knowledge* and *low self-efficacy*). Convincing people that simple at-home body weight exercises and not gym memberships could be helpful was also challenging (*low motivation*). All of this informed a logic model of the problem from the target population's perspective. Subsequently, the planning group decided that the overall program goal was *to improve the self-management of nonspecific LBP to reduce pain-related disability*. The personal determinants were knowledge and awareness, skills, fear avoidance and catastrophizing, self-efficacy, and motivation and outcome expectations.

Step 2: Program Outcomes and Objectives—Logic Model of Change

In this step, we created a matrix specifying the intended behavioral change expected to result from selfBACK intervention. The result was a schematic representation of what patients with LBP needed to do to reach the overall objective of self-managing their LBP, as exemplified in Table 1.

Table 1. Extract of the matrix of change objectives for the following behavioral outcome: To increase use of evidence-based self-management strategies. The full matrix is available in Multimedia Appendix 1 (Table S1).

Behavioral outcome[a]	Personal determinants				
	Knowledge and awareness[b]	Skills[c]	Fear avoidance and catastrophizing[d]	Self-efficacy[e]	Motivation and outcome expectations[f]
PO[g] 1: Accept self-management as treatment strategy for LBP[h] and make the decision to self-manage LBP with support from selfBACK app					
Change objectives	• Identify positive characteristics of self-management and negative characteristics of provider dependent behavior • List examples of self-management of LBP	Demonstrate ability to operate selfBACK app	• Recognize fearful thoughts and negative thinking in relation to self-management	Express confidence in the ability to operate selfBACK app	• Express positive feelings or thoughts about engaging in self-management of LBP • Expect that self-managing will ease living with LBP and achieving life goals
PO[g] 14: Integrate self-management strategies for LBP[h] into daily life					
Change objectives	• List ways to integrate self-management of LBP into daily routines	Demonstrate ability to schedule self-management into daily routines	• Recognize fearful and negative thoughts and feelings in relation to integrating self-management into daily routines • Recognize own fear-avoidance behavior in relation to integrating self-management into daily routines	Express confidence in the ability to integrate self-management of LBP into daily routines	• Express positive feelings or thoughts about integrating self-management into daily routines • Expect that integration of LBP self-management will lead to a healthier, better life

[a]Increase use of evidence-based self-management strategies.

[b]Increase knowledge of self-management behavior.

[c]Develop ability to engage in self-management behavior.

[d]Reduce fear or negative expectancies about engaging in self-management behavior.

[e]Improve perceived ability to uptake and engage in self-management behavior.

[f]Improve autonomous motivation to engage in self-management behavior and improve expectations to the outcome of self-management behavior.

[g]PO: performance objective.

[h]LBP: low back pain.

Step 3: Program Design

Core Components

Using evidence for the 3 main intervention components from clinical guidelines [5-8,28-31] and studies on management of LBP [32], as well as patient leaflets from the National Institute for Health and Care Excellence [33-41] and LBP management websites [42-45], three overall themes emerged: physical exercise, education, and physical activity.

Physical Exercise

Guidelines de campo [28,46] and systematic reviews on LBP treatment [47-53] endorse physical exercises for the management of persistent LBP. There is no evidence that any one type of exercise is better than others; however, strength training and motor control exercises are most commonly used [54,55]. *Exercises for flexibility* that aim to restore or improve the range of motion of the lumbar spine are also often part of programs to alleviate LBP [47,48,53]. In addition, reviews on motor control exercise support another type of exercise, *pain relief*, as being effective to support management of pain [47,50,52,53,56]. These exercises comprised movements performed in the midrange and without strong muscle contractions to facilitate controlled and smooth movements.

The amount of exercise was not always clearly described in RCTs; however, longer durations of exercise periods and heavier training seemed to be more effective in reducing back pain compared with shorter exercise periods and lighter loads [57]. The American College of Sports Medicine recommends 2 to 3 weekly sessions for muscle training at 60%-70% of 1 repetition

maximum for beginners and 80% or 1 repetition maximum, sets of 8 to 12 repetitions for strength and power, and >15 for endurance for those who are more experienced [58]. To maintain a good range of motion, flexibility exercises to the end range are recommended 2 to 3 days a week, held for 30 seconds, and repeated 2 to 4 times [58]. As the literature does not clearly indicate the most effective exercises, we chose to design individually tailored exercise programs based on, for example, symptoms, preferences or fitness levels, and exercises aimed at pain control or pain reduction, improved motor control, strength, and flexibility. Consequently, exercises included in the selfBACK exercise bank were categorized into six different targets: (1) flexibility exercises, (2) pain-relieving exercises in addition to strength exercises, (3) back extensors, (4) gluteal and hip muscles, (5) abdominal muscles, and (6) core muscles. The organization of exercises by their target, rules for progression, or regression between exercise levels were guided by consensus discussions among experienced clinicians and researchers, physiological reasoning, and clinical experience.

Education

A cognitive behavioral approach to patient education revolving around teaching or promoting pain coping skills such as activity pacing and progression guidance, goal setting and action planning, and mindfulness techniques [30], as well as reassurance about the prognosis of LBP [31], was recommended by the guidelines [28,29,31]. Furthermore, we chose education that comprised information about the condition, consequences and management, discouragement of bed rest, and advice to stay physically active. Our systematic reviews, conducted in step 1, also highlighted the importance of providing support that was easy for a user to integrate within the many competing activities of their daily lives [11,22]. Themes in relation to education, which were extracted from the guidelines [5-8,28-30], patient leaflets [33-41], and pain management websites [42-44], were organized under the main and subthemes, as presented in Textbox 1.

Textbox 1. Main themes and subthemes of the educational content.

First aid for acute back pain

- First aid reassurance
- First aid stay active

Fitting self-management into your daily life

- Daily activities
- Me time

General information about low back pain

- Cause of low back pain
- Guidelines low back pain
- Imaging
- Pain rating
- Reassurance
- Start exercise
- Stay active
- Structure of back

How body and mind are connected

- Mind–body connection

How thoughts, behavior, attitudes, and feelings affect low back pain

- Accepting pain
- Anxious thoughts and feelings
- Attitude
- Changing negative thoughts
- Distraction
- Distress
- Fear avoidance
- Stress
- Thoughts

How to overcome barriers for self-management of low back pain

- Barrier facilities
- Barrier family and work
- Barrier time
- Barrier tiredness
- Barrier support
- Barrier weather

How to practice mindfulness for low back pain

- Mindfulness

How to reach your goals

- Pacing

How to seek social support

- Family and friends

- Work

How to set specific, measurable, achievable, realistic, and timely (SMART) goals

- Action planning
- Goal setting

How to sleep better at night

- Sleep tips

How to solve problems

- Problem solving

Low back pain and other medical conditions

- Anxiety
- Depression
- Musculoskeletal pain
- Sleep problems

Self-manage your low back pain

- Encouragement to self-management

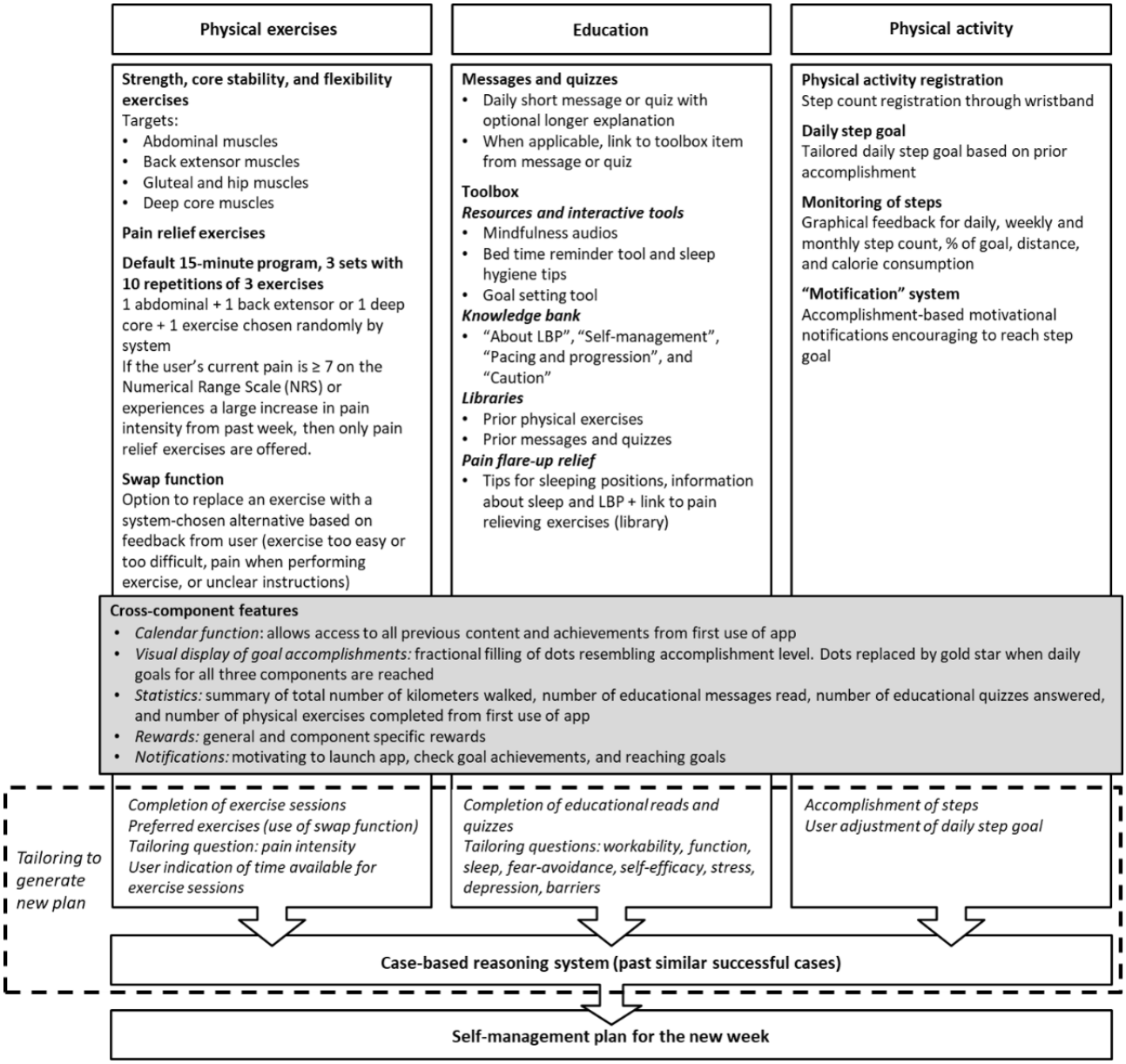
Physical Activity

The importance of physical activity has been recognized as a prime strategy in guidelines for self-management of LBP, including advice to stay active and at work, as well as discouragement of bed rest [5-8,28-30]. This has, to a large part, been based on the deconditioning model of LBP [59]; that is, patients with LBP may be restricted in the performance of everyday physical activities. Consequently, they risk developing an inactive lifestyle, and a vicious circle may then gradually develop. The recommended daily step count was based on the general recommendations for physical activity [58]. A minimum step count goal of 3000 per day was chosen to reflect the fact that participants may have functional disabilities that affect their physical activity level. Optimally, users should reach 10,000 steps per day [58,60].

Outline of Intervention App Design

The blueprint of the selfBACK intervention content that is incorporated in the selfBACK app is shown in Figure 1. Weekly SMPs with content from the three main intervention components are outlined as follows: (1) a bank of physical exercises; (2) a bank of educational messages and quizzes; and (3) physical activity registration in terms of step counts, together with an accomplishment-based motivational notification system. The bank of physical exercises has 59 strength and flexibility exercises organized in 5 targets with up to 6 difficulty levels and 11 additional pain relief exercises. Instructions for each exercise were provided in text and video formats, with real-life models demonstrating the exercises (no audio). Exercise sessions were recommended to be performed 3 to 5 times a week. The default program comprised 3 exercises constituting 15 minutes per session. An exercise could be swapped with another chosen by the system based on the participants' reason for swapping (eg, exercise too difficult or easy, pain when performing, or unclear instructions). The educational material comprised daily messages or quizzes and a toolbox. Messages and quizzes were structured under the 14 main themes with up to 9 subthemes identified in step 2 (Textbox 1). Short messages (<140 characters) were followed by an optional long message (maximum 500 characters), which included a more thorough explanation. Quizzes with yes or no answers were followed by the correct answer and an explanation. When applicable, the messages were accompanied with a link to a toolbox item with additional content to support self-management, that is, interactive tools (goal setting or bedtime reminder), mindfulness audios, and an explanatory text about how LBP, self-management, and the selfBACK app were connected. Additional toolbox items included 2 libraries with all previous physical exercises and educational messages or quizzes and advice on handling LBP relapse. Physical activity (step count) was tracked using a wearable device (Mi Band 3, Xiaomi Corp) connected to the app via Bluetooth. Daily, weekly, and monthly accomplishment of steps were graphically available to the participant. On the basis of their accomplishments, motivational notifications (*motifications*) encouraged participants to reach their daily step count goal.

Figure 1. Outline of the selfBACK intervention. Tailoring questions presented in Sandal et al [20]. LBP: low back pain.



Furthermore, the app had five cross-component features:

- A calendar function allowed users to access content and see accomplishments from the first day of using the app.
- Dots that gradually filled up as the daily goals were reached displayed the accomplishment level. On days where all 3 goals were reached, a gold star replaced the dots.
- A statistical summary showed the distance walked and number of messages read, quizzes answered, and exercises performed since the start of using the app.
- General and component-specific rewards were unlocked once the users performed the goal. For example, reaching all daily goals for the first time gave a reward. The next levels were 3, 7, and 14 times; another example was the number of steps in total, where the first reward was at 50,000 steps, followed by 100,000, 200,000, and 800,000 steps.
- Notifications motivated users to engage with the app but could be turned off by the user. Examples of these *motifications* were "Fantastic performance today! Achieved your step count goal!" and "You're more than half way to your step-count goal. Taking the stairs instead of the lift can really help towards your step count goal."

Tailoring of SMPs to individual participants was based on data from four sources: (1) the baseline questionnaire [20]; (2) the participant's achievement of physical exercises in the preceding week; (3) the participant's achievement of steps in the preceding week; and (4) weekly tailoring of question and answer sessions with a changing selection of questions from the baseline questionnaire [20], the participant's indication of time available for physical exercise per session, the participant's preference for exercises, and the participant's adjustment of the step goal for the new week (–10% to +10% to 20% from what the app suggested based on the preceding week), with an upper limit of 10,000 steps. A CBR system supported by a sophisticated rule engine used the abovementioned data from the current participant (case) and previous similar participants (cases) to generate new SMPs [19,25]. Generation of a new SMP was initiated by the weekly tailoring session or upon return to the

app if the participant had not used the app for more than a week. This process is described in detail elsewhere [19,25].

Underpinning Theories for Behavior Change and Engagement in DHIs

Self-management interventions can be characterized as behavior change interventions in that they are designed to help the patient learn and adopt a set of health behaviors and thus, benefit their condition [61]. To underpin the selfBACK intervention theoretically in relation to behavior change and engagement, 2 frameworks were applied in the design. The Transtheoretical Domain Framework (TDF) is an overall framework encompassing 33 behavior change theories [62]. TDF has been validated for use in behavior change and implementation research [63]. To facilitate the application of BCTs [64,65], matrices have been created by expert consensus, mapping BCTs and theoretical constructs as per TDF [66,67]. Furthermore, NPT underpins the strategies for uptake and adherence to the intervention. NPT is a sociological theory that has been widely used to understand the factors that influence how technologies or therapies are implemented, embedded, and integrated into daily routines [68-70]. NPT has four main constructs: (1) coherence—the sense-making work that participants undertake, which influences whether they are willing to embed a new practice in their lives; (2) cognitive participation—the work that participants undertake to engage with the new practice; (3) collective action—the work that participants do to enact a new practice; and (4) reflexive monitoring—the appraisal work that participants undertake to determine whether the new practice is worth sustaining or how it must be reconfigured to fit their needs [68-70]. All practical applications of the change objectives are linked to the BCTs and NPT domains, as exemplified in Table 2.

Table 2. Example matrix for mapping practical applications of performance and change objectives to BCTs[a] and NPT[b] domains. The full matrix is available in Multimedia Appendix 1 (Table S2).

Personal determinants and change objectives	Practical application	BCTs as per BCT taxonomy version 1 [64]	NPT domains [68,69]
PO[c] 1: accept self-management as a treatment strategy for LBP[d] and make the decision to self-manage LBP with support from selfBACK app			
• Knowledge and awareness: identify positive characteristics of self-management and negative characteristics of provider-dependent behavior and list examples of self-management of LBP • Skills: demonstrate the ability to operate selfBACK app • Fear avoidance and catastrophizing: recognize fearful thoughts and negative thinking in relation to self-management • Self-efficacy: express confidence in the ability to operate selfBACK app • Motivation and outcome expectations: express positive feelings or thoughts about engaging in self-management of LBP and expect that self-managing will ease living with LBP and achieving life goals	• Introduction session explaining structure and content of app, automatically shown after first log-in and thereafter accessible from Settings • Educational messages and quizzes • Referral from educational messages to relevant toolbox elements • Toolbox elements: resources and interactive tools, knowledge bank, libraries, and pain flare-up relief • Visual display of goal accomplishments • Rewards for achievements • Calendar function • Statistics • Notifications	• 5.1 Information about health consequences • 5.3 Information about emotional consequences • 4.1. Instruction on how to perform the behavior • 2.2. Feedback on behavior • 10.4. Social reward • 7.1. Prompts/cues • 15.1. Verbal persuasion about capability	• Coherence (gaining an understanding of the condition) • Collective action (developing skills) • Cognitive participation (engaging with the user to promote uptake) • Reflexive monitoring (evaluation and feedback)
PO[c] 14: integrate self-management strategies for LBP[d] into daily life			
• Knowledge and awareness: list ways to integrate self-management of LBP into daily routines • Skills: demonstrate the ability to schedule self-management into daily routines • Fear avoidance and catastrophizing: recognize fearful and negative thoughts and feelings in relation to integrating self-management into daily routines and recognize own fear avoidance behavior in relation to integrating self-management into daily routines • Self-efficacy: express confidence in the ability to integrate self-management of LBP into daily routines • Motivation and outcome expectations.: express positive feelings or thoughts about integrating self-management into daily routines and expect that integration of LBP self-management will lead to a healthier, better life	• Educational messages and quizzes • Referral from educational messages to relevant toolbox elements • Toolbox elements: resources and interactive tools, knowledge bank, libraries, and pain flare-up relief • Visual display of goal accomplishments • Rewards for achievements • Calendar function • Statistics • Notifications • Physical activity registration • Monitoring of steps • Motification system	• 5.1. Information about health consequences • 5.3. Information about emotional consequences • 4.1. Instruction on how to perform the behavior • 1.5. Review behavioral goals • 1.6. Discrepancy between current behavior and goals • 10.4. Social reward • 2.2. Feedback on behavior • 10.5. Social incentive • 7.1. Prompts/cues • 15.1. Verbal persuasion about capability	• Coherence (understanding) • Cognitive participation (engaging with the user to promote uptake) • Reflexive monitoring (evaluation and feedback)

[a]BCT: behavior change technique.

[b]NPT: normalization process theory.

[c]PO: performance objective.

[d]LBP: low back pain.

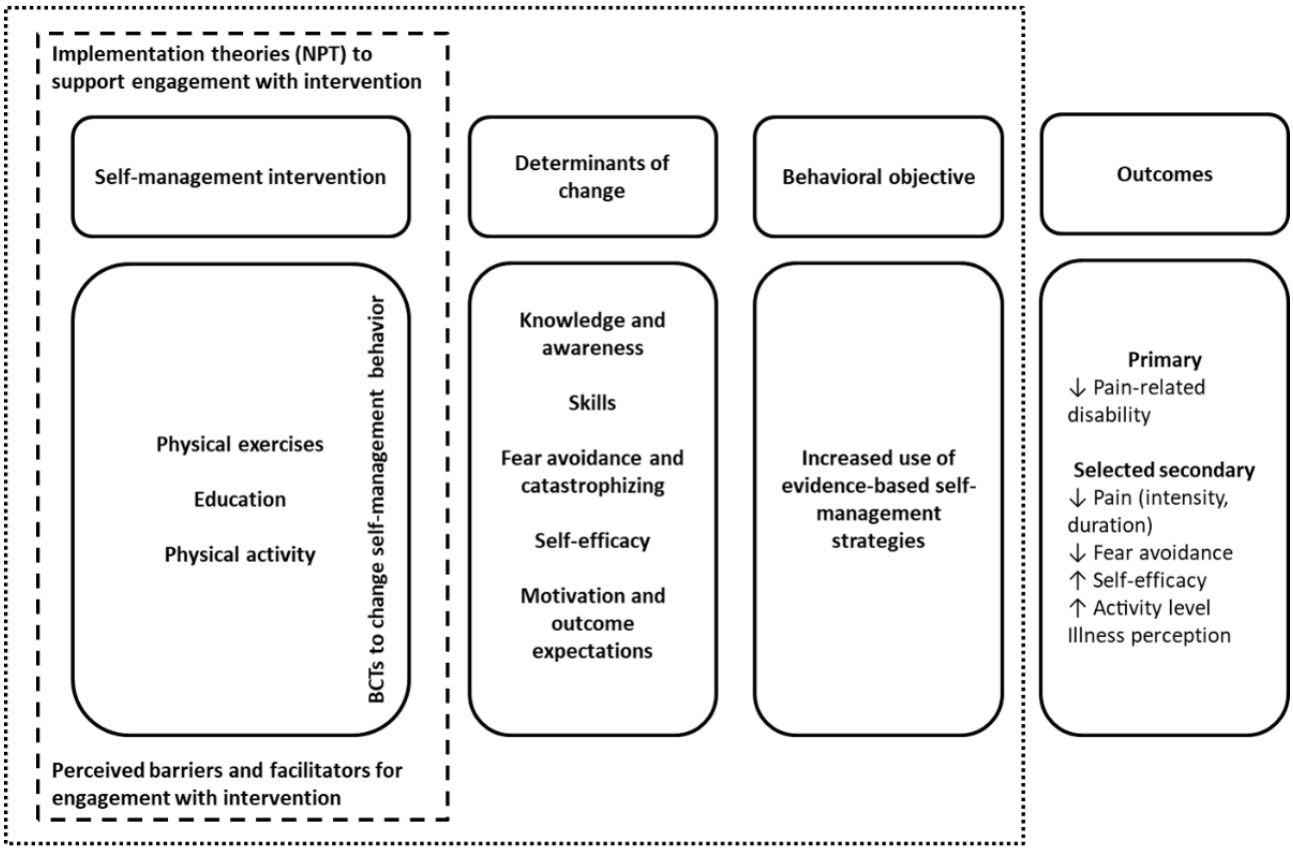
Outline of Engagement Strategies

Previous work on DHIs for self-management of LBP identified lack of engagement as a major barrier to use [13,71]. Machado et al [13] advocated the incorporation of strategies to increase engagement by stimulating repeated use, for example, through reminders, gamification, or reward systems. In our systematic review on barriers to and facilitators of engagement in DHIs, we further identified the briefness of information, feedback, tailoring, user-friendliness, design, and layout as facilitators of

enhanced engagement in self-management of LBP through DHIs [22]. Gamification, the concept of applying game mechanics to nongame contexts, has been shown to enhance user engagement in DHIs by using many different techniques, for example, badges, progress elements, quizzes, and challenges [72,73]. In addition, gamification in DHIs offers other advantages such as enhancing motivation, making health activities enjoyable and understandable, and improving users' abilities to self-manage their condition [73]. Therefore, it is thought to contribute to behavior change because of its resemblance with established health BCTs [74]. Our interviews with people with LBP confirmed the empirical findings; a friendly, supportive tone in the app that motivated to do more rather than pointing out insufficiencies was suggested to improve engagement, as well as records of accomplishment and rewards. Feedback on self-management behavior was thought to increase motivation and engagement. Appraisal of the app by the HCP or visibility of a trustworthy source was also important for participants.

A program logic model bringing together the intended effect outcomes, behavioral outcomes, targeted determinants, and program outputs is presented in Figure 2. This intervention logic model describes the mechanistic pathway from the intervention to the reduction in pain-related disability because of LBP.

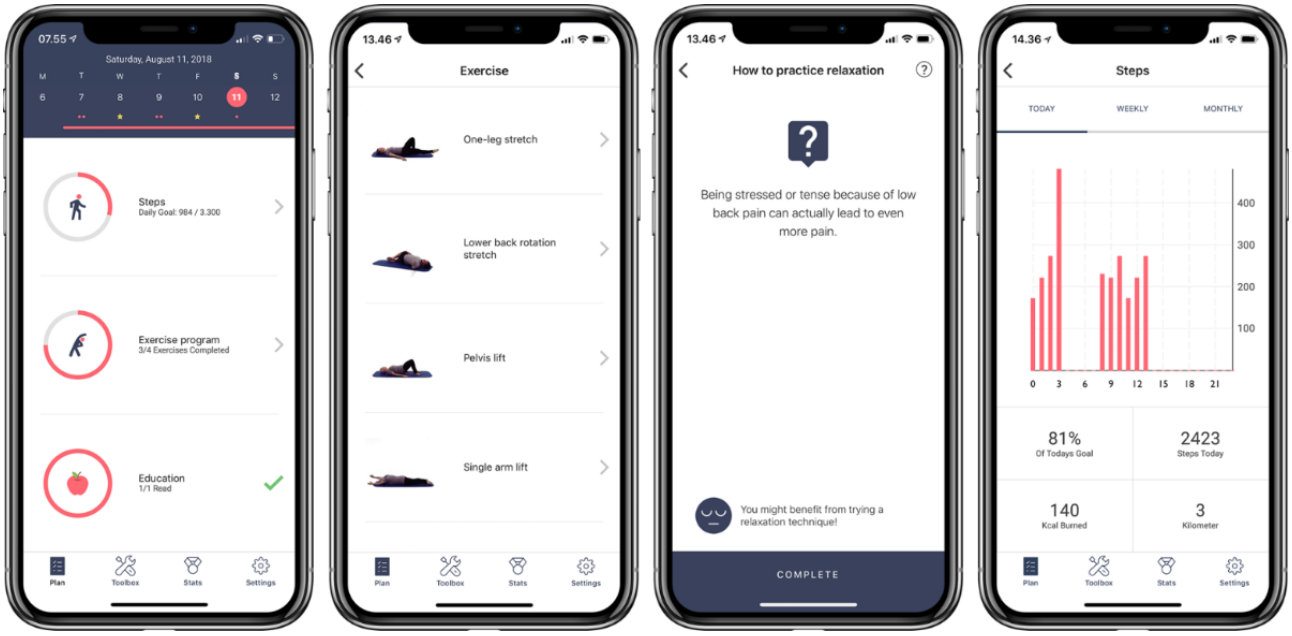
Figure 2. Program logic model of the selfBACK intervention. BCT: behavior change technique; NPT: normalization process theory.



Step 4: Program Production

Results from the user journey map tests and continuous user testing of app beta versions provided insights into user needs and potential barriers to use and contributed to improvements in the intervention design and layout (Figure 3). A full description of the app design process and technical specifications, including the personas and user journey map, frontend development, coding, and security measures, are available elsewhere or upon requesting the authors [75].

Figure 3. Screenshot of the selfBACK app plan screen showing the 3 main components of a weekly self-management plan and a screenshot of each of the three main components of the self-management intervention: physical exercises, education, and physical activity.



Results from the feasibility studies improved the activity monitoring from the wristband to the app; established a notification taxonomy taking both achievements and reported barriers into account; and refined the educational content, step count accuracy and goal setting, motivational notifications, and technical barriers to using the app [26]. Finally, the 6-week single-arm pilot study in Denmark and Norway resulted in refinements of recruitment, screening, and app installation procedures as well as effect and process evaluations for the RCT [27].

Step 5: Implementation Plan

The result of the fifth step was a plan of how we foresaw adoption and implementation in the selfBACK RCT (Table 3).

The program use outcomes (adoption and implementation) were formulated with eight performance objectives. Changeable determinants as per NPT [68,69] were identified and crossed with the performance objectives to formulate change objectives. BCTs formed the theoretical foundation for the methods that underpinned the practical strategies for adoption and implementation [64]. A screening and inclusion procedure that lies between the adoption and implementation outcomes has been published [20,27]. On the basis of the results of the pilot study, practical strategies for adoption were refined, including an upscale in the number of recruitment sites or HCPs needed to adopt the intervention to reach the desired number of participants in the RCT.

Table 3. Plan for program adoption and implementation of the selfBACK intervention.

Program use outcomes and performance objectives	Determinants for embedment of digital intervention to everyday routine as per NPT[a] [71,72]	Change objectives	BCTs[b] to address each change objective as per BCT taxonomy version 1 [67]	Practical strategies
Adoption use outcome: Recruitment sites (HCPs[c]) adopt the selfBACK intervention and participant recruitment procedures				
Agree to participate in selfBACK RCT[d] as recruitment sites	Coherence and cognitive participation	• Managers at recruitment sites provide verbal agreement to allow their service to implement selfBACK • Managers at recruitment sites allocate resources (time) to support recruitment	• 1.8. Behavioral contract	• Verbal agreement from each recruitment site manager to allow their clinic to implement selfBACK as an add-on to usual care and support participant recruitment to selfBACK RCT[d] • Recruitment site managers to nominate HCPs to receive instruction in recruitment pathway from selfBACK researchers
Agree to recruit patients to participate in selfBACK RCT	Cognitive participation	• HCPs develop an understanding of the purpose, structure, and content of the selfBACK intervention	• 1.8. Behavioral contract	• HCPs receive information about selfBACK RCT and receive information from selfBACK researchers if there are further questions
HCPs implement recruitment procedures (identification of eligible patients based on RCT inclusion criteria, including practical strategies for establishing contact between patient and selfBACK research team)	Coherence and cognitive participation	• HCPs develop an understanding of who eligible patients are • HCPs develop skills to initiate recruitment	• 5.1. Information about health consequences of the intervention • 4.1. Instruction on how to perform a behavior	• HCPs receive information about selfBACK RCT and receive information from selfBACK researchers if there are further questions • HCPs receive recruitment pathway ideas from selfBACK researchers
HCPs encourage patients to participate in selfBACK RCT as an add-on to usual care	Cognitive participation and collective action	• HCPs inform eligible patients about the selfBACK intervention	• 4.1. Instruction on how to perform a behavior	• HCPs give written or verbal information about selfBACK to patients and promote participation verbally
Implementation use outcome: participants engage in the selfBACK intervention by implementing selfBACK app in daily routines				

Program use outcomes and performance objectives	Determinants for embedment of digital intervention to everyday routine as per NPT[a] [71,72]	Change objectives	BCTs[b] to address each change objective as per BCT taxonomy version 1 [67]	Practical strategies
Participants make sense of the selfBACK intervention	Coherence	• Participants develop understanding of the purpose and potential of the selfBACK app	• 8.1. Behavioral practice/rehearsal	• Participants explore the features of the selfBACK app[e]
Participants build and sustain engagement in the selfBACK intervention	Cognitive participation	• Participants initiate regular use of the selfBACK app	• 8.1. Behavioral practice/rehearsal	• Participants launch and use the selfBACK app regularly[e]
Participants invest efforts and resources in engagement in the selfBACK intervention	Collective action	• Participants prioritize regular use of the selfBACK app	• 8.1. Behavioral practice/rehearsal	• Participants launch and use the selfBACK app regularly[e]
Participants evaluate engagement in the selfBACK intervention	Reflexive monitoring	• Participants appraise the selfBACK app and decide to sustain engagement	• 5.1. Information about health consequences • 5.6. Information about emotional consequences	• Participants deem selfBACK app as effective or helpful[e] • Participants answer follow-up questionnaires during the trial period

[a]NPT: normalization process theory.

[b]BCT: behavioral change technique.

[c]HCP: health care professional.

[d]RCT: randomized controlled trial.

[e]Practical application is the desired scenario.

Discussion

Principal Findings

This study provides a detailed example of using IM to systematically develop a theory and evidence-based app-based intervention for people with nonspecific LBP. There is currently limited literature on the development of complex digital interventions for self-management, and this study should inform future researchers in this evolving field.

During our IM process, we found that the existing knowledge on self-management of LBP is generic with respect to descriptions of intervention content. Clinical practice guidelines for LBP also provide incomplete descriptions of the advice delivered in interventions [76]. In our systematic review, a consistent finding was that a comprehensive description of the development and use of theory was either brief or absent from the included studies [11]. A recent study revealed that there are >700 pain-related smartphone apps available from the main app stores and 61 apps about LBP solely [13]. However, few of these apps have evidence-based content, many have not been rigorously tested for effectiveness, and HCPs and patients are rarely involved in their development [13]. In contrast to the existing apps, selfBACK was developed with a strong theoretical underpinning, involved key stakeholders, integrated feedback from users, and was developed by experts in relevant fields.

As demonstrated in this study, the IM approach details how accessing and using theory can support the development of intervention content as requested in the Medical Research Council's framework [15] and the CONSORT-EHEALTH checklist [16]. A thorough description of an intervention's theoretical underpinning increases the understanding of the mechanisms of action and the potential for replication. The application of the IM approach with strong theoretical underpinning will allow for meaningful evaluation of the process of implementing digital self-management of LBP in both people's daily lives and primary care.

Strengths and Limitations

A strength of this study is the >2-year development phase following IM. A large and diverse planning group worked continuously on developing, testing, and refining the selfBACK intervention, which made the uniqueness of combining behavior change theory and implementation theory, empirical findings, and state-of-the-art technical solutions such as CBR possible. Rigorous pretesting, an essential part of IM [14], with 2 feasibility studies, a pilot study, and continual user testing involving >200 people, is also a strength of this study. A possible weakness of our study is that we did not fully adhere to the IM protocol. Our needs assessment and the resulting logic model of the problem encountered self-management of LBP exclusively from the target population's perspective. The IM protocol advocates a multilevel approach, taking environmental factors and the potential influence of environmental agents into account [14].

Conclusions

This paper reports a detailed example of the application of IM in the development of a theory-driven complex DHI designed to support self-management for people with LBP. Although IM is a time-intensive collaborative process, this report of the range of methods used provides a transparent account of the development process and a blueprint for designing and creating future DHIs for self-management.

Acknowledgments

The selfBACK project has received funding from the European Union Horizon 2020 Research and Innovation Program under grant agreement 689043. The authors wish to thank all selfBACK consortium members for their efforts in developing the intervention.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Matrix of change objectives for the behavioral outcome “To increase use of evidence-based self-management strategies” and Mapping practical applications of performance and change objectives to behavior change techniques and normalization process theory domains.
[DOCX File , 47 KB - jmir_v24i1e26555_app1.docx]

References

1. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, et al. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2163-2196 [FREE Full text] [doi: 10.1016/S0140-6736(12)61729-2] [Medline: 23245607]
2. Murray CJ, Vos T, Lozano R, Naghavi M, Flaxman AD, Michaud C, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012 Dec 15;380(9859):2197-2223. [doi: 10.1016/S0140-6736(12)61689-4] [Medline: 23245608]
3. Lidgren L. The bone and joint decade 2000-2010. Bull World Health Organ 2003;81(9):629 [FREE Full text] [Medline: 14710501]
4. Buchbinder R, Blyth FM, March LM, Brooks P, Woolf AD, Hoy DG. Placing the global burden of low back pain in context. Best Pract Res Clin Rheumatol 2013 Oct;27(5):575-589. [doi: 10.1016/j.berh.2013.10.007] [Medline: 24315140]
5. van Tulder M, Becker A, Bekkering T, Breen A, del Real MT, Hutchinson A, COST B13 Working Group on Guidelines for the Management of Acute Low Back Pain in Primary Care. Chapter 3. European guidelines for the management of acute nonspecific low back pain in primary care. Eur Spine J 2006 Mar;15 Suppl 2:S169-S191 [FREE Full text] [doi: 10.1007/s00586-006-1071-2] [Medline: 16550447]
6. Airaksinen O, Brox JI, Cedraschi C, Hildebrandt J, Klaber-Moffett J, Kovacs F, COST B13 Working Group on Guidelines for Chronic Low Back Pain. Chapter 4. European guidelines for the management of chronic nonspecific low back pain. Eur Spine J 2006 Mar;15 Suppl 2:S192-S300 [FREE Full text] [doi: 10.1007/s00586-006-1072-1] [Medline: 16550448]
7. Delitto A, George SZ, Van Dillen L, Whitman JM, Sowa G, Shekelle P, Orthopaedic Section of the American Physical Therapy Association. Low back pain. J Orthop Sports Phys Ther 2012 Apr;42(4):A1-57 [FREE Full text] [doi: 10.2519/jospt.2012.42.4.A1] [Medline: 22466247]
8. Savigny P, Watson P, Underwood M, Guideline Development Group. Early management of persistent non-specific low back pain: summary of NICE guidance. BMJ 2009 Jun 04;338:b1805. [doi: 10.1136/bmj.b1805] [Medline: 19502217]
9. Koes BW, van Tulder M, Lin CC, Macedo LG, McAuley J, Maher C. An updated overview of clinical guidelines for the management of non-specific low back pain in primary care. Eur Spine J 2010 Dec;19(12):2075-2094 [FREE Full text] [doi: 10.1007/s00586-010-1502-y] [Medline: 20602122]
10. Murray E, Burns J, See TS, Lai R, Nazareth I. Interactive Health Communication Applications for people with chronic disease. Cochrane Database Syst Rev 2005 Oct 19(4):CD004274. [doi: 10.1002/14651858.CD004274.pub4] [Medline: 16235356]
11. Nicholl BI, Sandal LF, Stochkendahl MJ, McCallum M, Suresh N, Vasseljen O, et al. Digital support interventions for the self-management of low back pain: a systematic review. J Med Internet Res 2017 May 21;19(5):e179 [FREE Full text] [doi: 10.2196/jmir.7290] [Medline: 28550009]
12. Garg S, Garg D, Turin TC, Chowdhury MF. Web-based interventions for chronic back pain: a systematic review. J Med Internet Res 2016 Jul 26;18(7):e139 [FREE Full text] [doi: 10.2196/jmir.4932] [Medline: 27460413]
13. Machado GC, Pinheiro MB, Lee H, Ahmed OH, Hendrick P, Williams C, et al. Smartphone apps for the self-management of low back pain: a systematic review. Best Pract Res Clin Rheumatol 2016 Dec;30(6):1098-1109. [doi: 10.1016/j.berh.2017.04.002] [Medline: 29103552]

14. Bartholomew L, Markham C, Ruiter R, Fernández M, Kok G, Parcel G. Planning Health Promotion Programs: An Intervention Mapping Approach. 4th edition. Hoboken, New Jersey: Wiley; 2016.
15. Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. Int J Nurs Stud 2013 May;50(5):587-592. [doi: 10.1016/j.ijnurstu.2012.09.010] [Medline: 23159157]
16. Eysenbach G, CONSORT-EHEALTH Group. CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. J Med Internet Res 2011 Dec 31;13(4):e126 [FREE Full text] [doi: 10.2196/jmir.1923] [Medline: 22209829]
17. Hurley DA, Murphy LC, Hayes D, Hall AM, Toomey E, McDonough SM, et al. Using intervention mapping to develop a theory-driven, group-based complex intervention to support self-management of osteoarthritis and low back pain (SOLAS). Implement Sci 2016 Apr 26;11:56 [FREE Full text] [doi: 10.1186/s13012-016-0418-2] [Medline: 27113575]
18. Detaille SI, van der Gulden JW, Engels JA, Heerkens YF, van Dijk FJ. Using intervention mapping (IM) to develop a self-management programme for employees with a chronic disease in the Netherlands. BMC Public Health 2010 Jun 21;10:353 [FREE Full text] [doi: 10.1186/1471-2458-10-353] [Medline: 20565925]
19. Mork PJ, Bach K, selfBACK Consortium. A decision support system to enhance self-management of low back pain: protocol for the selfBACK project. JMIR Res Protoc 2018 Jul 20;7(7):e167 [FREE Full text] [doi: 10.2196/resprot.9379] [Medline: 30030208]
20. Sandal LF, Stochkendahl MJ, Svendsen MJ, Wood K, Øverås CK, Nordstoga AL, et al. An app-delivered self-management program for people with low back pain: protocol for the selfBACK randomized controlled trial. JMIR Res Protoc 2019 Dec 03;8(12):e14720 [FREE Full text] [doi: 10.2196/14720] [Medline: 31793897]
21. Rasmussen CD, Svendsen MJ, Wood K, Nicholl BI, Mair FS, Sandal LF, et al. App-delivered self-management intervention trial selfBACK for people with low back pain: protocol for implementation and process evaluation. JMIR Res Protoc 2020 Oct 29;9(10):e20308 [FREE Full text] [doi: 10.2196/20308] [Medline: 33118959]
22. Svendsen MJ, Wood KW, Kyle J, Cooper K, Rasmussen CD, Sandal LF, et al. Barriers and facilitators to patient uptake and utilisation of digital interventions for the self-management of low back pain: a systematic review of qualitative studies. BMJ Open 2020 Dec 12;10(12):e038800 [FREE Full text] [doi: 10.1136/bmjopen-2020-038800] [Medline: 33310794]
23. Kjaer P, Kongsted A, Ris I, Abbott A, Rasmussen CD, Roos EM, et al. GLA:D Back group-based patient education integrated with exercises to support self-management of back pain - development, theories and scientific evidence. BMC Musculoskelet Disord 2018 Nov 29;19(1):418 [FREE Full text] [doi: 10.1186/s12891-018-2334-x] [Medline: 30497440]
24. SELFBACK—activity recognition for self-management of low back pain. In: Research and Development in Intelligent Systems XXXIII. Cham: Springer; 2016.
25. Bach K, Szczepanski T, Aamodt A, Gundersen O, Mork P. Case representation and similarity assessment in the selfBACK decision support system. In: Proceedings of the International Conference on Case-Based Reasoning. 2016 Presented at: International Conference on Case-Based Reasoning; Oct 31 - Nov 2, 2016; Atlanta, Georgia. [doi: 10.1007/978-3-319-47096-2_3]
26. Nordstoga AL, Bach K, Sani S, Wiratunga N, Mork PJ, Villumsen M, et al. Usability and acceptability of an app (SELFBACK) to support self-management of low back pain: mixed methods study. JMIR Rehabil Assist Technol 2020 Sep 09;7(2):e18729 [FREE Full text] [doi: 10.2196/18729] [Medline: 32902393]
27. Sandal LF, Øverås CK, Nordstoga AL, Wood K, Bach K, Hartvigsen J, et al. A digital decision support system (selfBACK) for improved self-management of low back pain: a pilot study with 6-week follow-up. Pilot Feasibility Stud 2020 May 23;6:72 [FREE Full text] [doi: 10.1186/s40814-020-00604-2] [Medline: 32489674]
28. de Campos TF. Low back pain and sciatica in over 16s: assessment and management NICE Guideline [NG59]. J Physiother 2017 Apr;63(2):120 [FREE Full text] [doi: 10.1016/j.jphys.2017.02.012] [Medline: 28325480]
29. Qaseem A, Wilt TJ, McLean RM, Forciea MA, Clinical Guidelines Committee of the American College of Physicians, Denberg TD, et al. Noninvasive treatments for acute, subacute, and chronic low back pain: a clinical practice guideline from the american college of physicians. Ann Intern Med 2017 Apr 04;166(7):514-530 [FREE Full text] [doi: 10.7326/M16-2367] [Medline: 28192789]
30. Wong JJ, Côté P, Sutton DA, Randhawa K, Yu H, Varatharajan S, et al. Clinical practice guidelines for the noninvasive management of low back pain: a systematic review by the Ontario Protocol for Traffic Injury Management (OPTIMa) Collaboration. Eur J Pain 2017 Feb;21(2):201-216. [doi: 10.1002/ejp.931] [Medline: 27712027]
31. Oliveira CB, Maher CG, Pinto RZ, Traeger AC, Lin CC, Chenot J, et al. Clinical practice guidelines for the management of non-specific low back pain in primary care: an updated overview. Eur Spine J 2018 Nov;27(11):2791-2803. [doi: 10.1007/s00586-018-5673-2] [Medline: 29971708]
32. Maher C, Underwood M, Buchbinder R. Non-specific low back pain. The Lancet 2017 Feb;389(10070):736-747 [FREE Full text] [doi: 10.1016/s0140-6736(16)30970-9]
33. Low back pain and sciatica in over 16s: assessment and management. National Institute for Health and Care Excellence. 2016. URL: https://www.nice.org.uk/guidance/ng59/ifp/chapter/Low-back-pain-and-sciatica-the-care-you-should-expect [accessed 2022-01-07]

34. Low back pain - information for patients. Oxford University Hospitals NHS Trust. 2013. URL: https://www.ouh.nhs.uk/patient-guide/leaflets/files/5712Plowbackpain.pdf [accessed 2022-01-07]
35. Section 10, Chapter 3: self-management of low back pain. International Society for the Study of the Lumbar Spine. URL: https://www.wheelessonline.com/issls/section-10-chapter-3-self-management-of-low-back-pain/ [accessed 2022-01-07]
36. Roland M, Waddell G, Klaber-Moffett J, Burton K, Main C. The Back Book: The Best Way to Deal With Back Pain. London: TSO; 2002.
37. What you should know about your chronic low back pain. Institute of Health Economics. 2015. URL: https://actt.albertadoctors.org/CPGs/Lists/CPGDocumentList/LBP-Patient-Handout-Chronic.pdf [accessed 2022-01-07]
38. What you should know about your acute low back pain. Institute of Health Economics. 2015. URL: https://actt.albertadoctors.org/CPGs/Lists/CPGDocumentList/LBP-Patient-Handout-Acute.pdf [accessed 2022-01-07]
39. Acute Low Back Pain: So your back hurts... Learn what works, what doesn't, and how to help yourself. Institute for Work & Health. 2015. URL: https://www.iwh.on.ca/sites/iwh/files/iwh/tools/so_your_back_hurts_2010.pdf [accessed 2022-01-07]
40. Chronic Low Back Pain: So your back hurts... Learn what works, what doesn't, and how to help yourself. The Institute for Work & Health. 2015. URL: https://www.iwh.on.ca/sites/iwh/files/iwh/tools/so_your_back_hurts_2010.pdf [accessed 2022-01-07]
41. Advice for managing low back pain. Orthopaedic Spinal Services. 2011. URL: https://www.sahealth.sa.gov.au/wps/wcm/connect/a61c510049e4d938b3aefb3a89b74631/ManagingLowBackPain-RAH-AlliedHealth-120123.pdf?MOD=AJPERES&CACHEID=ROOTWORKSPACE-a61c510049e4d938b3aefb3a89b74631-n5iNhYW [accessed 2022-01-07]
42. PainACTION. Back Pain.: National Institutes of Health URL: https://www.painaction.com/back-pain/ [accessed 2017-04-19]
43. Your Guide to Pain Management: A Road Map for PainACTION.com, 2nd Edition. Newton, MA: Inflexxion; 2012.
44. Self-managing chronic pain. The Pharmacy Guild of Australia. URL: https://www.painaustralia.org.au/static/uploads/files/painaust-factsheet10-final-wfbdhlqkosbm.pdf [accessed 2022-01-07]
45. Pain management - Learn how to help manage your musculoskeletal pain through clinically supported tips, tools, research and more. PainHEALTH. URL: https://painhealth.csse.uwa.edu.au/pain-management/ [accessed 2022-01-07]
46. Pillastrini P, Gardenghi I, Bonetti F, Capra F, Guccione A, Mugnai R, et al. An updated overview of clinical guidelines for chronic low back pain management in primary care. Joint Bone Spine 2012 Mar;79(2):176-185. [doi: 10.1016/j.jbspin.2011.03.019] [Medline: 21565540]
47. van Middelkoop M, Rubinstein SM, Verhagen AP, Ostelo RW, Koes BW, van Tulder MW. Exercise therapy for chronic nonspecific low-back pain. Best Pract Res Clin Rheumatol 2010 Apr;24(2):193-204. [doi: 10.1016/j.berh.2010.01.002] [Medline: 20227641]
48. Searle A, Spink M, Ho A, Chuter V. Exercise interventions for the treatment of chronic low back pain: a systematic review and meta-analysis of randomised controlled trials. Clin Rehabil 2015 Dec;29(12):1155-1167. [doi: 10.1177/0269215515570379] [Medline: 25681408]
49. Gordon R, Bloxham S. A systematic review of the effects of exercise and physical activity on non-specific chronic low back pain. Healthcare (Basel) 2016 Apr 25;4(2):22 [FREE Full text] [doi: 10.3390/healthcare4020022] [Medline: 27417610]
50. Steffens D, Maher CG, Pereira LS, Stevens ML, Oliveira VC, Chapple M, et al. Prevention of low back pain: a systematic review and meta-analysis. JAMA Intern Med 2016 Feb;176(2):199-208. [doi: 10.1001/jamainternmed.2015.7431] [Medline: 26752509]
51. Saragiotto BT, Maher CG, Yamato TP, Costa LO, Costa LC, Ostelo RW, et al. Motor control exercise for nonspecific low back pain: a cochrane review. Spine (Phila Pa 1976) 2016 Aug 15;41(16):1284-1295. [doi: 10.1097/BRS.0000000000001645] [Medline: 27128390]
52. Macedo LG, Saragiotto BT, Yamato TP, Costa LO, Menezes Costa LC, Ostelo RW, et al. Motor control exercise for acute non-specific low back pain. Cochrane Database Syst Rev 2016 Feb 10;2:CD012085. [doi: 10.1002/14651858.CD012085] [Medline: 26863390]
53. Hayden JA, van Tulder MW, Malmivaara A, Koes BW. Exercise therapy for treatment of non-specific low back pain. Cochrane Database Syst Rev 2005 Jul 20(3):CD000335. [doi: 10.1002/14651858.CD000335.pub2] [Medline: 16034851]
54. van Tulder M, Malmivaara A, Esmail R, Koes B. Exercise therapy for low back pain: a systematic review within the framework of the cochrane collaboration back review group. Spine (Phila Pa 1976) 2000 Nov 01;25(21):2784-2796. [doi: 10.1097/00007632-200011010-00011] [Medline: 11064524]
55. Babatunde OO, Jordan JL, Van der Windt DA, Hill JC, Foster NE, Protheroe J. Effective treatment options for musculoskeletal pain in primary care: a systematic overview of current evidence. PLoS One 2017;12(6):e0178621 [FREE Full text] [doi: 10.1371/journal.pone.0178621] [Medline: 28640822]
56. Gomes-Neto M, Lopes JM, Conceição CS, Araujo A, Brasileiro A, Sousa C, et al. Stabilization exercise compared to general exercises or manual therapy for the management of low back pain: a systematic review and meta-analysis. Phys Ther Sport 2017 Jan;23:136-142. [doi: 10.1016/j.ptsp.2016.08.004] [Medline: 27707631]
57. Slade SC, Keating JL. Trunk-strengthening exercises for chronic low back pain: a systematic review. J Manipulative Physiol Ther 2006 Feb;29(2):163-173. [doi: 10.1016/j.jmpt.2005.12.011] [Medline: 16461178]

58. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, American College of Sports Medicine. American College of Sports Medicine position stand. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: guidance for prescribing exercise. Med Sci Sports Exerc 2011 Jul;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
59. Wittink H, Hoskins Michel T, Wagner A, Sukiennik A, Rogers W. Deconditioning in patients with chronic low back pain: fact or fiction? Spine (Phila Pa 1976) 2000 Sep 01;25(17):2221-2228. [doi: 10.1097/00007632-200009010-00013] [Medline: 10973406]
60. Di Iorio A, Abate M, Guralnik JM, Bandinelli S, Cecchi F, Cherubini A, et al. From chronic low back pain to disability, a multifactorial mediated pathway: the InCHIANTI study. Spine (Phila Pa 1976) 2007 Dec 15;32(26):E809-E815 [FREE Full text] [doi: 10.1097/BRS.0b013e31815cd422] [Medline: 18091475]
61. Michie S, van Stralen MM, West R. The behaviour change wheel: a new method for characterising and designing behaviour change interventions. Implement Sci 2011 Apr 23;6:42 [FREE Full text] [doi: 10.1186/1748-5908-6-42] [Medline: 21513547]
62. Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual Saf Health Care 2005 Feb;14(1):26-33 [FREE Full text] [doi: 10.1136/qshc.2004.011155] [Medline: 15692000]
63. Cane J, O'Connor D, Michie S. Validation of the theoretical domains framework for use in behaviour change and implementation research. Implement Sci 2012 Apr 24;7:37 [FREE Full text] [doi: 10.1186/1748-5908-7-37] [Medline: 22530986]
64. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013 Aug;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
65. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
66. Michie S, Atkins L, West R. The Behaviour Change Wheel : A Guide to Designing Interventions. United Kingdom: Silverback; 2014.
67. Michie S, Johnston M, Francis J, Hardeman W, Eccles M. From theory to intervention: mapping theoretically derived behavioural determinants to Behaviour Change Techniques. Applied Psychol 2008 Oct;57(4):660-680. [doi: 10.1111/j.1464-0597.2008.00341.x]
68. May C, Finch T. Implementing, embedding, and integrating practices: an outline of normalization process theory. Sociol 2009 Jun 15;43(3):535-554. [doi: 10.1177/0038038509103208]
69. May CR, Cummings A, Girling M, Bracher M, Mair FS, May CM, et al. Using Normalization Process Theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. Implement Sci 2018 Jun 07;13(1):80 [FREE Full text] [doi: 10.1186/s13012-018-0758-1] [Medline: 29879986]
70. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
71. Platt A, Outlay C, Sarkar P, Karnes S. Evaluating user needs in wellness apps. Int J Human Comput Interact 2015 Oct 22;32(2):119-131. [doi: 10.1080/10447318.2015.1099803]
72. Sardi L, Idri A, Fernández-Alemán JL. A systematic review of gamification in e-Health. J Biomed Inform 2017 Jul;71:31-48 [FREE Full text] [doi: 10.1016/j.jbi.2017.05.011] [Medline: 28536062]
73. Looyestyn J, Kernot J, Boshoff K, Ryan J, Edney S, Maher C. Does gamification increase engagement with online programs? A systematic review. PLoS One 2017;12(3):e0173403 [FREE Full text] [doi: 10.1371/journal.pone.0173403] [Medline: 28362821]
74. Edwards EA, Lumsden J, Rivas C, Steed L, Edwards LA, Thiyagarajan A, et al. Gamification for health promotion: systematic review of behaviour change techniques in smartphone apps. BMJ Open 2016 Oct 04;6(10):e012447 [FREE Full text] [doi: 10.1136/bmjopen-2016-012447] [Medline: 27707829]
75. Deliverable D4.6: demonstration of the final connected SELFBACK mobile application (DEMCON) on iOS and Android: SELFBACK - A decision support system for self-management of low back pain. Southern Denmark University. URL: https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5bbdc8e12&appId=PPGMS [accessed 2022-01-07]
76. Stevens ML, Lin CC, de Carvalho FA, Phan K, Koes B, Maher CG. Advice for acute low back pain: a comparison of what research supports and what guidelines recommend. Spine J 2017 Oct;17(10):1537-1546. [doi: 10.1016/j.spinee.2017.05.030] [Medline: 28713052]

Abbreviations

BCT: behavior change technique

CBR: case-based reasoning
CONSORT-EHEALTH: Consolidated Standards of Reporting Trials of Electronic and Mobile Health Applications and Online Telehealth
DHI: digital health intervention
HCP: health care professional
IM: intervention mapping
LBP: low back pain
NPT: normalization process theory
RCT: randomized controlled trial
SMP: self-management plan
TDF: Transtheoretical Domains Framework

Edited by G Eysenbach; submitted 16.12.20; peer-reviewed by J Ayre, J Parsons; comments to author 19.02.21; revised version received 07.03.21; accepted 04.11.21; published 24.01.22.

Please cite as:
Svendsen MJ, Sandal LF, Kjær P, Nicholl BI, Cooper K, Mair F, Hartvigsen J, Stochkendahl MJ, Søgaard K, Mork PJ, Rasmussen C
Using Intervention Mapping to Develop a Decision Support System–Based Smartphone App (selfBACK) to Support Self-management of Nonspecific Low Back Pain: Development and Usability Study
J Med Internet Res 2022;24(1):e26555
URL: https://www.jmir.org/2022/1/e26555
doi:10.2196/26555
PMID:35072645

Original Paper

Effectiveness of Internet-Based Cognitive Behavioral Therapy With Telephone Support for Noncardiac Chest Pain: Randomized Controlled Trial

Terje Thesen[1,2], MD; Joseph A Himle[3], PhD; Egil W Martinsen[4†], PhD; Liv T Walseth[1], PhD; Frode Thorup[5], MD; Frode Gallefoss[6,7], PhD; Egil Jonsbu[2,8], PhD

[1]Distriktspsykiatrisk senter Solvang, Sørlandet Hospital, Sørlandet sykehus Helse Foretak, Kristiansand, Norway
[2]Faculty of Medicine and Health Science, Norwegian University of Science and Technology, Trondheim, Norway
[3]School of Social Work and School of Medicine-Psychiatry, University of Michigan, Ann Arbor, MI, United States
[4]Institute of Clinical Medicine, University of Oslo, Oslo, Norway
[5]Department of Cardiology, Sørlandet Hospital, Sørlandet sykehus Helse Foretak, Kristiansand, Norway
[6]Faculty of Medicine, University of Bergen, Bergen, Norway
[7]Department of Clinical Research, Sørlandet Hospital, Sørlandet sykehus Helse Foretak, Kristiansand, Norway
[8]Department of Psychiatry, Møre and Romsdal Hospital Trust, Molde, Norway
[†]deceased

Corresponding Author:
Terje Thesen, MD
Distriktspsykiatrisk senter Solvang
Sørlandet Hospital
Sørlandet sykehus Helse Foretak
Servicebox 416
Kristiansand, 4604
Norway
Phone: 47 38174800
Email: terje.thesen@sshf.no

Abstract

Background: Noncardiac chest pain has a high prevalence and is associated with reduced quality of life, anxiety, avoidance of physical activity, and high societal costs. There is a lack of an effective, low-cost, easy to distribute intervention to assist patients with noncardiac chest pain.

Objective: In this study, we aimed to investigate the effectiveness of internet-based cognitive behavioral therapy with telephone support for noncardiac chest pain.

Methods: We conducted a randomized controlled trial, with a 12-month follow-up period, to compare internet-based cognitive behavioral therapy to a control condition (treatment as usual). A total of 162 participants aged 18 to 70 years with a diagnosis of noncardiac chest pain were randomized to either internet-based cognitive behavioral therapy (n=81) or treatment as usual (n=81). The participants in the experimental condition received 6 weekly sessions of internet-based cognitive behavioral therapy. The sessions covered different topics related to coping with noncardiac chest pain (education about the heart, physical activity, interpretations/attention, physical reactions to stress, optional panic treatment, and maintaining change). Between sessions, the participants also engaged in individually tailored physical exercises with increasing intensity. In addition to internet-based cognitive behavioral therapy sessions, participants received a brief weekly call from a clinician to provide support, encourage adherence, and provide access to the next session. Participants in the treatment-as-usual group received standard care for their noncardiac chest pain without any restrictions. Primary outcomes were cardiac anxiety, measured with the Cardiac Anxiety Questionnaire, and fear of bodily sensations, measured with the Body Sensations Questionnaire. Secondary outcomes were depression, measured using the Patient Health Questionnaire; health-related quality of life, measured using the EuroQol visual analog scale; and level of physical activity, assessed with self-report question. Additionally, a subgroup analysis of participants with depressive symptoms at baseline (PHQ-9 score ≥5) was conducted. Assessments were conducted at baseline, posttreatment, and at 3- and 12-month follow-ups. Linear mixed models were used to evaluate treatment effects. Cohen *d* was used to calculate effect sizes.

Results: In the main intention-to-treat analysis at the 12-month follow-up time point, participants in the internet-based cognitive behavioral therapy group had significant improvements in cardiac anxiety (–3.4 points, 95% CI –5.7 to –1.1; *P*=.004, *d*=0.38) and a nonsignificant improvement in fear of bodily sensations (–2.7 points, 95% CI –5.6 to 0.3; *P*=.07) compared with the treatment-as-usual group. Health-related quality of life at the 12-month follow-up improved with statistical and clinical significance in the internet-based cognitive behavioral therapy group (8.8 points, 95% CI 2.8 to 14.8; *P*=.004, *d*=0.48) compared with the treatment-as-usual group. Physical activity had significantly (*P*<.001) increased during the 6-week intervention period for the internet-based cognitive behavioral therapy group. Depression significantly improved posttreatment (*P*=.003) and at the 3-month follow-up (*P*=.03), but not at the 12-month follow-up (*P*=.35). Participants with depressive symptoms at baseline seemed to have increased effect of the intervention on cardiac anxiety (*d*=0.55) and health-related quality of life (*d*=0.71) at the 12-month follow-up. In the internet-based cognitive behavioral therapy group, 84% of the participants (68/81) completed at least 5 of the 6 sessions.

Conclusions: This study provides evidence that internet-based cognitive behavioral therapy with minimal therapist contact and a focus on physical activity is effective in reducing cardiac anxiety and increasing health related quality of life in patients with noncardiac chest pain.

Trial Registration: ClinicalTrials.gov NCT03096925; http://clinicaltrials.gov/ct2/show/NCT03096925

(J Med Internet Res 2022;24(1):e33631) doi:10.2196/33631

KEYWORDS

noncardiac chest pain; internet-based treatment; internet-assisted treatment; cognitive behavioral therapy; psychosomatic medicine; randomized controlled trial; pain; treatment; internet-based cognitive behavioral therapy; effectiveness; support; intervention

Introduction

Background

Noncardiac chest pain is the most common final diagnosis for patients presenting with chest pain at a cardiac unit [1]. Almost half of patients with noncardiac chest pain experience ongoing complaints that negatively affect their overall quality of life after initial noncardiac chest pain diagnosis, and the associated societal costs are also high [2,3]. Typical consequences include cardiac anxiety, fear of bodily sensations, and avoidance of physical activity [2,4,5].

Psychological treatment based on cognitive behavioral therapy, delivered face-to-face by trained clinicians, has yielded promising results for noncardiac chest pain [6]. There are, however, several challenges with the dissemination of ordinary face-to-face cognitive behavioral therapy to patients with noncardiac chest pain, including a shortage of specialists who provide cognitive behavioral therapy, lack of motivation among patients with noncardiac chest pain to receive psychological treatment, and the fact that typical care providers (eg, cardiologists and cardiac nurses) often do not have the skills to provide such treatment [6].

Evidence-based face-to-face cognitive behavioral therapy for noncardiac chest pain is quite time-consuming and costly, with a normal treatment span of 6 to 12 sessions, each lasting 45 to 60 minutes each. To address this barrier, 2 randomized controlled trials [7,8] tested shorter duration face-to-face cognitive behavioral therapy specifically designed for noncardiac chest pain. For patients experiencing sustained noncardiac chest pain–related complaints 6 months after diagnosis, a 3-session face-to-face cognitive behavioral therapy with emphasis on exposure to physical activity was effective in reducing noncardiac chest pain–related complaints [7]. Furthermore, depression was a significant predictor of poor outcome [2], and fear of bodily sensations was an important mediator for change [7]. Conversely, another large randomized controlled trial [8], which included an unselected group of patients with noncardiac chest pain immediately after receiving their noncardiac chest pain diagnosis, found that 3 to 4 sessions of face-to-face cognitive behavioral therapy were not effective in reducing noncardiac chest pain–related complaints. It is difficult to fully determine the reasons for these disparate results; however, one [7] focused on physical activity and included patients who had significant complaints 6 months after the initial diagnosis of noncardiac chest pain; in contrast, the other [8] did not specifically focus on physical activity and included all patients with noncardiac chest pain immediately after their cardiac evaluation, regardless of the severity of their symptoms and duration of illness.

Internet-based cognitive behavioral therapy has the potential to increase accessibility and can be delivered at reduced costs compared with those of face-to-face cognitive behavioral therapy. Two pilot studies [9,10] have shown promising results for noncardiac chest pain: a small randomized controlled trial [9] (intervention group: n=7; treatment-as-usual control group: n=8) found that a 4-session internet-based cognitive behavioral therapy yielded greater reductions in cardiac anxiety and depression; however, because it was designed to reveal probable differences between the groups, it was underpowered to test statistical significance. Mourad et al [9] concluded that their brief internet-based cognitive behavioral therapy intervention was feasible to deliver, and though it was likely to be effective, a larger sample would be needed to adequately evaluate effectiveness. Concurrently and independently, our research group found that 6-session internet-based cognitive behavioral therapy was feasible and led to a significant reduction in cardiac anxiety for unselected noncardiac chest pain patients in an uncontrolled, open pilot study (n=10) [10]. Participants in this study were offered the internet-based cognitive behavioral therapy immediately after a cardiac condition had been excluded and received a brief weekly therapist call in addition to discuss

home assignments and reinforce session content. The intervention content was based on a brief face-to-face treatment previously found to be effective in [7], and it had a similar emphasis on physical activity.

Objectives

The purpose of this study was to further evaluate the effectiveness of the internet-based cognitive behavioral therapy intervention for noncardiac chest pain from our pilot study [10], in a randomized controlled trial with sufficient statistical power. The primary goal was to investigate the effects of the intervention on cardiac anxiety and fear of bodily sensations. The secondary goals were to evaluate changes in depression, quality of life, and physical activity, and to investigate the differential effects of this internet-based cognitive behavioral therapy in a subgroup of participants with depressive symptoms at baseline.

Methods

Design

This study was a 2-arm randomized controlled trial. One arm received treatment as usual, and the other received internet-based cognitive behavioral therapy for noncardiac chest pain. Participants were assessed pre- and posttreatment as well as at 3- and 12-month follow-up time points.

Participants and Recruitment

Participants were recruited at the Sørlandet Hospital, Kristiansand, Norway. A member of the research group screened all patients with chest pain as their main complaint who had been referred to the cardiac unit (including an inpatient and an outpatient unit) for participation. The member of the research group informed cardiologists about possible eligible patients. The cardiologists used a checklist with inclusion and exclusion criteria as well as the definition of noncardiac chest pain to determine the eligibility of each patient. The cardiologist briefly explained the project to eligible patients. If the patient agreed, a member of the research group immediately provided further verbal and written information about the trial before the recruitment procedure.

Eligible patients were aged 18 to 70 years, had no cardiac or other somatic disease that could explain their chest pain symptoms, no history of or ongoing severe heart disease, and did not meet any of the following exclusion criteria: (1) language difficulties; (2) inability to perform at least moderate physical activity due to physical constraints; (3) obvious cognitive impairment (eg, severe intellectual disability, psychosis, dementia, or intoxication); (4) no regular access to a computer or tablet with internet connection; and (5) severe somatic comorbidities (eg, cancer, severe kidney failure).

If patients had been examined with coronary computed tomography angiography, only patients with coronary artery blockage less than 50% were eligible. All eligible patients were asked to sign a paper consent form if they agreed to participate.

Intervention

The brief internet-based cognitive behavioral therapy with minimal therapist contact has been previously described [10]. The intervention was adjusted and improved based on feedback from participants in the pilot study [10]. Internet-based cognitive behavioral therapy sessions (Figure 1; Multimedia Appendix 1) were completed autonomously once per week for 6 weeks. Between sessions, participants received support calls from a therapist.

The internet-based cognitive behavioral therapy was specifically designed and adapted for noncardiac chest pain (1) to provide an alternative explanation for their chest pain, (2) to provide tools and education on how to handle bodily discomfort, (3) to encourage physical activity with the aim that the participants experienced its safety, and (4) to provide information on strategies that can prevent relapse.

The first session was completed on a computer or tablet at the cardiac unit immediately after the cardiac examination was completed. The sessions covered different topics relevant to the noncardiac chest pain patients (Figure 2). At the end of every session, each participant completed a detailed activity plan for the following week. Adherence to these activities was reported to the therapist electronically. The activity plan focused on individually tailored physical exercises with increasing intensity. In addition, participants were encouraged to perform an attention task regularly after the third session. Once a week, a brief support call (5-7 minutes duration), designed to reinforce the session content, discuss home-based tasks, and troubleshoot problems, was made by the same therapist. The support call had 3 predefined elements: (1) discussion about the previous week's physical activities and tasks, (2) discussion about the previous week's session, and (3) general information about the following week's session. The participants were given access to the subsequent week's internet-based cognitive behavioral therapy session after the support call. The completed sessions remained accessible to the participant throughout the project period.

Participants in the control group received treatment as usual—personal consultation with a doctor at the cardiac unit after completing the cardiologic investigations. The consultation included information regarding the results of the examination and advice on medication, diet, and physical activity. The intervention group received internet-based cognitive behavioral therapy in addition to personal consultation at the cardiac unit. Lottery tickets (each with a price of €5, approximately US $5.67) were sent to all participants after they completed assessments to enhance adherence.

A unique code was used to access the intervention platform. The intervention platform included session-by-session content and homework exercises. Data were stored using pseudonyms on a secure server in Oslo, Norway. This server is backed up every 24 hours by another secure server in Amsterdam, Netherlands. Only the therapist could link the code to the participant. No personal identifiable information was collected, and IP addresses were not logged.

Figure 1. Screenshot from the internet-based cognitive behavioral therapy at week 3.

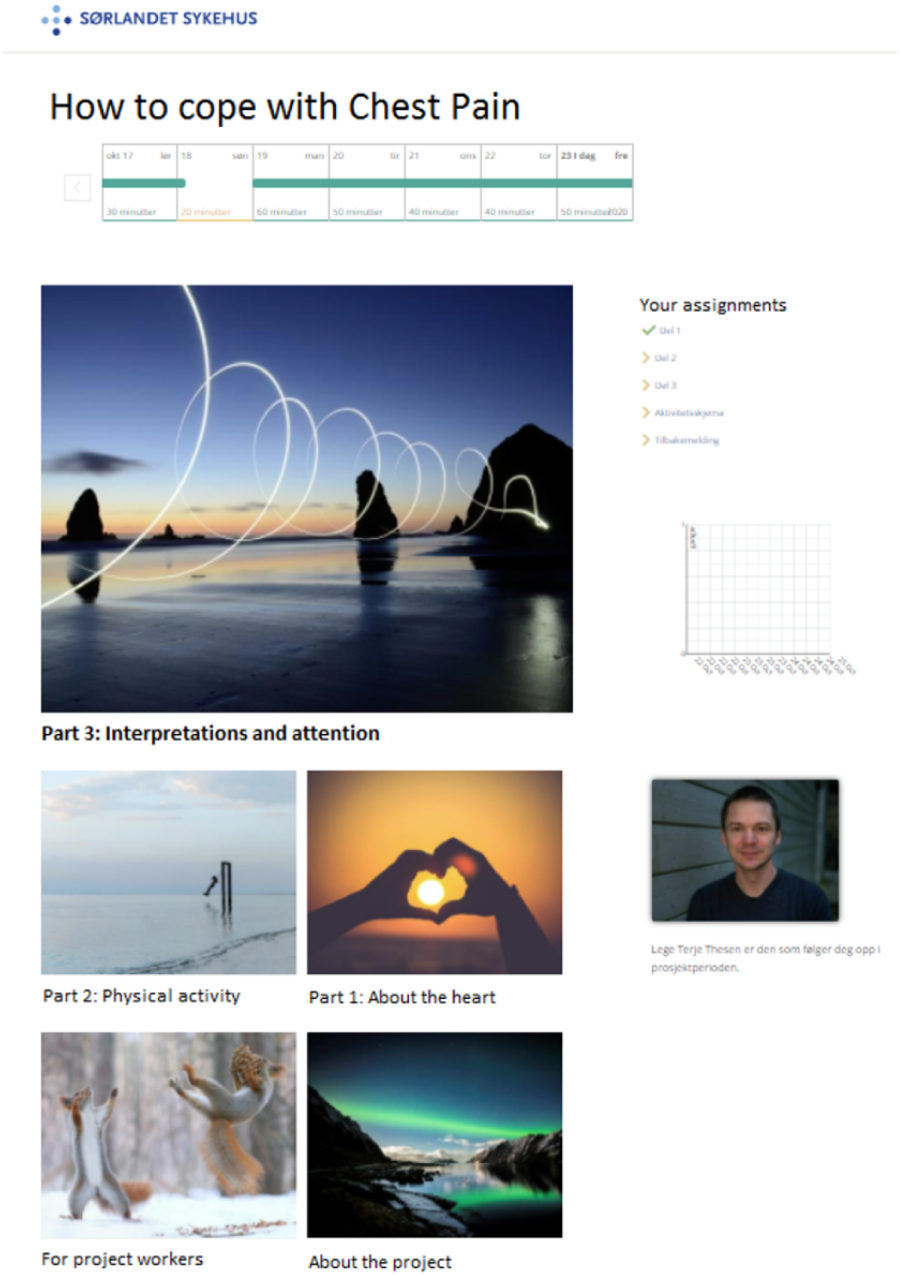
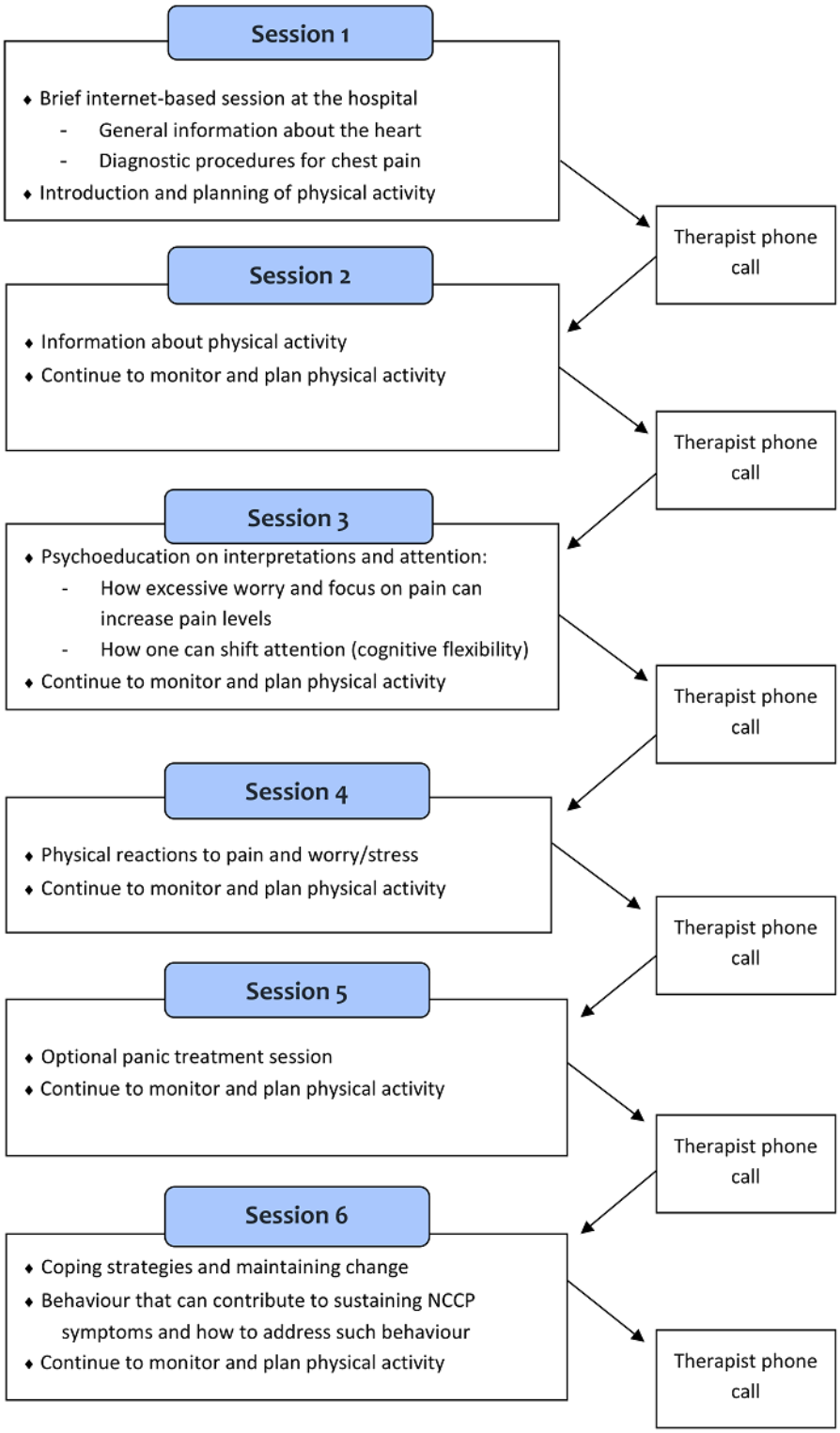


Figure 2. Flowchart for internet-based cognitive behavioral therapy. NCCP: noncardiac chest pain [10].



Assessments and Outcomes

Participants completed web-based self-report questionnaires at baseline and posttreatment, and at 3- and 12-month follow-up time points. The primary outcome measures were sum scores for the 18-item Cardiac Anxiety Questionnaire (CAQ), which measures worry or fear of heart sensations, avoidance of physical activity due to fear of eliciting heart symptoms, heart-focused attention, and reassurance seeking [11,12], and the 17-item Body Sensations Questionnaire (BSQ), which measures worry about specific bodily sensations [13].

The CAQ has a sum score that ranges from 0 to 72, with each item scored on a scale of 0 to 4. Higher scores indicate more symptoms. The CAQ is well validated, has good internal consistency (Cronbach α=.84), and high test–retest reliability (r=0.88) [14].

The BSQ has a total score that ranges from 17 to 85, with each item scored on a scale of 1 (not worried) to 5 (extremely frightened) [15]. The BSQ has been used in several noncardiac chest pain trials [4,7], has high internal consistency (Cronbach α=.87), and moderate test–retest reliability (r=0.66) [13].

Secondary outcome measures were the 9-item Patient Health Questionnaire (PHQ-9) score, the EuroQol Visual Analog Scale (EQ-VAS) score, and physical activity assessment.

The PHQ-9 includes and assesses level of depression. Each item is scored from 0 to 3, and the sum range ranges from 0 to 27. Higher scores indicate greater depression. We used a sum score ≥5 to indicate mild depressive symptoms. The PHQ-9 is well validated, has high internal consistency (Cronbach α=.89), high test–retest reliability (r=0.84) [16], and also validated in a web-based format [17].

The EQ-VAS measures the user's overall rating of their health [18] using a visual analog scale that is presented as a line from 0 to 100, where 0 is defined as the worst imaginable health state and 100 is the best.

Participants were further assessed using an investigator-developed, nonvalidated question that assessed physical activity level: "How many times each week on average do you perform physical activity more than 30 min?".

Clinical Relevance

No minimal clinically important difference thresholds have been established for the CAQ and BSQ. For PHQ-9, we defined a difference of 3 points or a reduction of 20% from baseline as the minimal clinically important difference based on values from studies [19-21]. The minimal clinically important difference for EQ-VAS has not been established specifically for noncardiac chest pain, but previous studies on cancer and COPD [22-24] have proposed that change in the range of 5.4-7.0 points is clinically relevant.

Adverse Events

The risk of serious adverse events due to internet-based cognitive behavioral therapy was considered low prior to the trial, and adverse events were not systematically collected. PHQ-9 scores were reviewed for each participant. Participants with a score ≥20 (severe depression) or a score of 3 on question 9 (which addresses suicidal ideation) were contacted. A decision by the principal investigator (TT) was made regarding the need for a psychiatric evaluation and whether to contact their general practitioner.

Randomization and Blinding

The participants were randomized 1:1 without stratification or the use of known blocks. The study used a web-based randomization procedure performed at a remote location (Web-CRF at the Norwegian University of Science and Technology in Trondheim). Because the intervention was internet-based cognitive behavioral therapy, blinding was not possible for the participant or therapist. Outcome measures were collected electronically, and except for the PHQ-9 scores, were not known to the therapist or the research group during the study period.

Sample Size

The study was powered to detect a Cohen d effect size of at least 0.5. The power calculations were based on the results from the pilot trial [10] and a previous study with brief face-to-face cognitive behavioral therapy for noncardiac chest pain [7]. A 2-sided test, with $\alpha<.05$ and power >80% was used, and a sample size of 63 participants in each group was needed to detect a mean difference between groups of 5.1 points on the BSQ with a standard deviation of 10.2. To tolerate an anticipated 20% dropout rate, a total sample size of 80 participants in each group was deemed appropriate.

Data Analysis

Primary analyses were conducted using intention-to-treat analysis, in which participants were analyzed according to the group they had been randomized. In the secondary per-protocol analysis, participants were excluded if they (1) did not complete at least 5 of 6 sessions; (2) did not complete all assessments; or (3) were found after randomization not to have been eligible based on the inclusion and exclusion criteria. Linear mixed models were used to evaluate treatment effects. Linear mixed models take into account repeated measures for each participant. The fixed part of the model included a group indicator variable (intervention or treatment as usual), a time variable (posttreatment, 3-month follow-up, and 12-month follow-up), and an interaction term between the group indicator and time variable. Baseline scores were included as covariates in the fixed portion of the analysis. Maximum likelihood estimations were used to estimate the effects on all outcomes. Between-group effect sizes (Cohen d) were calculated for selected time points. We used SPSS statistical software (version 25, IBM Corp) for all analyses.

Patient Involvement

One person with noncardiac chest pain experience participated in the development of the internet-based cognitive behavioral therapy intervention. This person reviewed each session during the development phase and provided input on the relevance of the topics and the didactics of the intervention.

Ethics

The study protocol was approved by the Regional Committee for Medical Research Ethics (2014/2031).

Results

Participants were recruited consecutively (Figure 3) from April 3, 2017 to March 26, 2018, with a 6-week pause from May to June 2017 (total recruitment period: 10.5 months).

Of 796 patients with chest pain who were screened for participation, 504 patients fulfilled criteria for noncardiac chest pain; 231 patients were invited to participate, and 162 accepted and were randomly allocated (intervention group: n=81; treatment-as-usual group: n=81). One participant withdrew consent and requested that their collected data be deleted; therefore, 161 participants were included in the intention-to-treat analysis (Table 1). One participant had severe comorbid disease, which was discovered after randomization, and another had severe depressive symptoms that required specialist treatment;

these patients were excluded from the per-protocol analysis. In the intervention group, 68 (84%) completed the internet-based cognitive behavioral therapy (at least 5 of 6 sessions). Posttreatment assessments were completed for 133, 3-month assessments for 126, and 12-month assessments for 117 participants. In the per-protocol analysis, 55 participants in the intervention group and 61 participants in the treatment-as-usual group were included. The groups were well balanced for demographic and clinical data.

Figure 3. Flow of participants during the study period. NCCP: non-cardiac chest pain; TAU: treatment as usual.

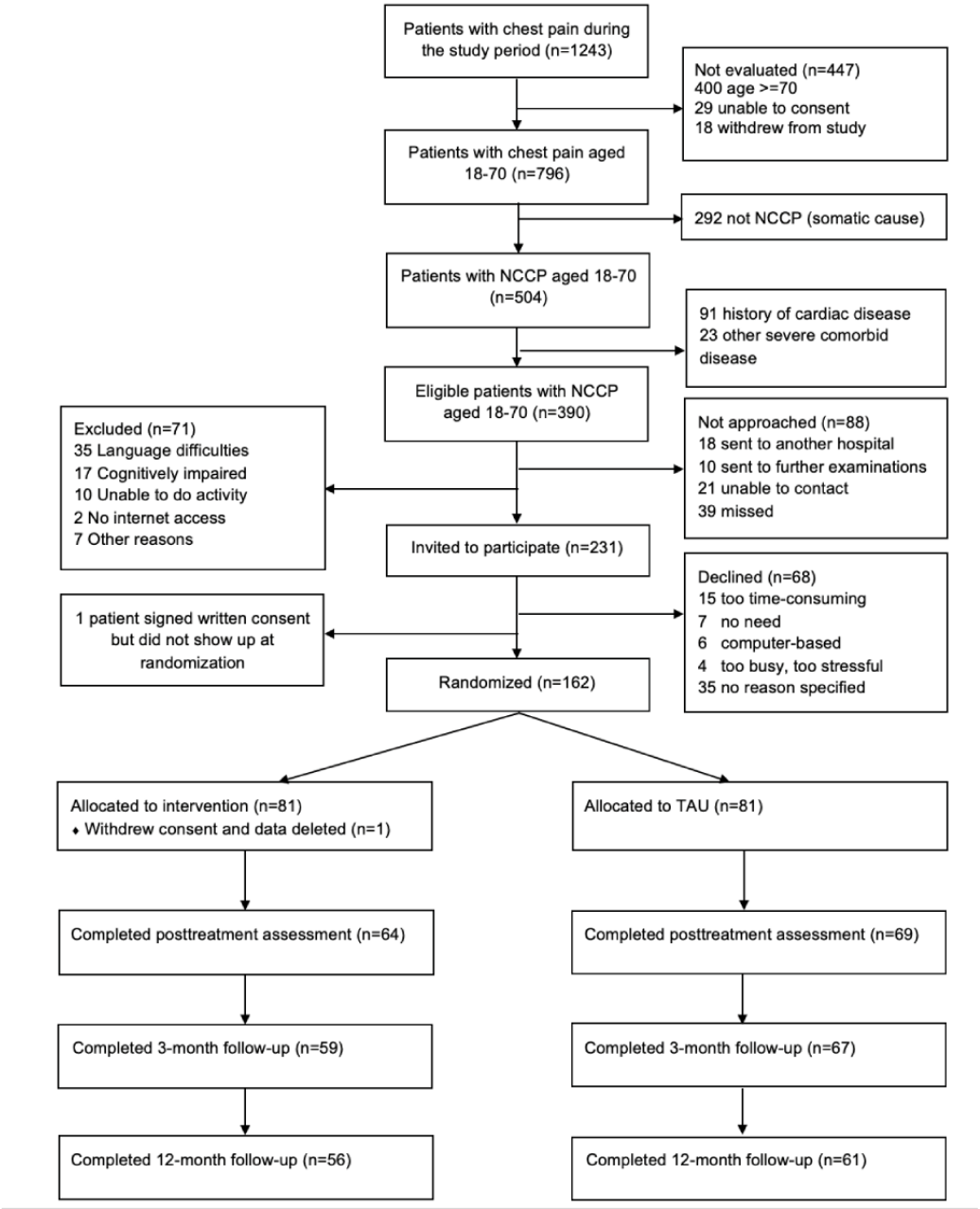


Table 1. Demographic and clinical data collected at baseline.

Characteristic	Intervention group (n=80), n (%)	Treatment as usual (n=81), n (%)	Full sample (n=161), n (%)
Gender			
Men	36 (45)	38 (47)	74 (46)
Women	44 (55)	43 (53)	87 (54)
Age (years), median (range)	53 (20-69)	51 (30-69)	52 (20-69)
Married or cohabiting	64 (80)	68 (84)	132 (82)
Highest education			
Primary or high school	18 (23)	18 (22)	36 (22)
Vocational school	29 (36)	30 (37)	59 (37)
College or university	33 (41)	33 (41)	66 (41)
Occupational status			
Full work	43 (54)	48 (59)	91 (57)
Part time work	16 (20)	9 (11)	25 (16)
Disability	14 (18)	18 (22)	32 (20)
Retired	4 (5)	5 (6)	9 (6)
Sick leave	3 (4)	1 (1)	4 (3)
Comorbid disorders[a]			
Somatic	42 (53)	41 (51)	82 (52)
Mental health	10 (13)	11 (14)	21 (13)
Chest pain duration			
0-1 month	22 (28)	24 (30)	46 (29)
2-6 months	29 (36)	28 (35)	57 (35)
>6 months	29 (36)	29 (36)	58 (36)
Chest pain frequency			
Never	2 (3)	4 (5)	6 (4)
Seldom	41 (51)	37 (46)	78 (48)
Weekly	20 (25)	26 (32)	46 (29)
Daily	17 (21)	14 (17)	31 (19)

[a]According to self-report.

Primary Outcomes

In the intention-to-treat analysis (Table 2), statistically significant reductions in CAQ score at the 3-month (P=.03) and 12-month follow-up (P=.004) were observed in the intervention group compared with the treatment-as-usual group. The effect size at 12-month follow-up was small to moderate (d=0.38). Accordingly, the BSQ scores were significantly reduced at posttreatment (P=.03) and at the 3-month follow-up (P=.04).

In the secondary per-protocol analysis (Table 2), participants in the intervention group showed statistically significant reductions in CAQ scores at the 3-month (P=.02) and 12-month (P=.002) follow-up compared with those in the treatment-as-usual group. Hence, the BSQ scores in the per-protocol analysis were significantly reduced at all time points (posttreatment: P=.02; 3-month follow-up: P=.008; 12-month follow-up: P=.04).

Table 2. Comparison of outcome measures between groups.

Outcome	Score, mean (SD)		Intention-to-treat analysis		Per-protocol analysis	
	Intervention[a]	Treatment as usual[b]	95% CI	*P* value[c]	95% CI	*P* value[d]
Primary						
Cardiac Anxiety Questionnaire						
Baseline	23.2 (9.8)	23.4 (10.3)	Reference	N/A[e]	Reference	N/A
Post	17.5 (9.1)	18.4 (9.7)	–1.0 (–3.3 to 1.2)	.36	–1.2 (–3.6 to 1.1)	.31
3 months	14.5 (7.9)	18.1 (10.4)	–2.5 (–4.8 to –0.2)	.03	–2.9 (–5.3 to –0.5)	.02
12 months	13.8 (8.7)	17.6 (11.1)	–3.4 (–5.7 to –1.1)	.004	–3.7 (–6.0 to –1.3)	.002
Body Sensations Questionnaire						
Baseline	36.8 (11.6)	36.7 (12.9)	Reference	N/A	Reference	N/A
Post	31.3 (9.7)	33.5 (12.5)	–3.1 (–6.0 to –0.3)	.03	–3.5 (–6.5 to –0.5)	.02
3 months	29.7 (9.9)	32.2 (12.6)	–3.0 (–5.9 to –0.1)	.04	–4.1 (–7.1 to –1.1)	.008
12 months	29.2 (10.3)	31.3 (12.2)	–2.7 (–5.6 to 0.3)	.07	–3.1 (–6.1 to –0.1)	.04
Secondary						
Patient Health Questionnaire–9						
Baseline	6.9 (4.7)	6.9 (5.0)	Reference	N/A	Reference	N/A
Post	4.9 (3.9)	6.6 (5.2)	–1.6 (–2.7 to –0.6)	.003	–2.1 (–3.2 to –0.9)	<.001
3 months	4.4 (3.5)	5.9 (4.6)	–1.3 (–2.4 to –0.2)	.03	–1.7 (–2.8 to –0.5)	.004
12 months	4.8 (3.8)	5.3 (5.0)	–0.5 (–1.7 to 0.6)	.35	–0.9 (–2.0 to 0.3)	.13
EuroQol Visual Analog Scale						
Baseline	63.5 (20.2)	63.4 (19.4)	Reference	N/A	Reference	N/A
Post	69.0 (19.9)	62.0 (21.6)	7.7 (1.9 to 13.4)	.009	10.5 (4.4 to 16.6)	.001
3 months	71.8 (19.0)	61.1 (20.6)	10.3 (4.5 to 16.2)	.001	12.1 (6.1 to 18.2)	<.001
12 months	71.8 (19.0)	61.8 (22.4)	8.8 (2.8 to 14.8)	.004	10.0 (4.0 to 16.1)	.001
Physical activity[f]						
Baseline	3.0 (1.9)	3.4 (2.6)	Reference	N/A	Reference	N/A
Post	4.5 (1.9)	3.3 (2.1)	1.3 (0.8 to 1.8)	<.001	1.6 (1.1 to 2.1)	<.001
3 months	3.7 (1.9)	3.4 (2.4)	0.3 (–0.2 to 0.8)	.19	0.5 (0 to 1.0)	.06
12 months	3.7 (2.0)	3.5 (2.1)	0.1 (–0.4 to 0.6)	.65	0.3 (–0.2 to 0.7)	.30

[a]Intervention: baseline (n=80); post (n=64); 3 months (n=59); 12 months (n=56).

[b]Treatment as usual: baseline (n=81); post (n=69); 3 months (n=67); 12 months (n=61).

[c]Estimated mean differences between groups adjusted for baseline values (linear mixed models).

[d]Estimated mean differences between groups adjusted for baseline values (linear mixed models); n=55 in the intervention group and n=61 in the treatment-as-usual group.

[e]N/A: not applicable.

[f]Number of times with moderate physical activity >30 minutes a week.

Secondary Outcomes

In the intention-to-treat analysis, participants in the intervention group reported a statistically significant improvement in depression posttreatment (*P*=.003) and at the 3-month follow-up (*P*=.03) compared with the treatment-as-usual group.

Health-related quality of life improved, reaching both statistical (posttreatment: *P*=.009; 3-month follow-up: *P*=.001; 12-month follow-up: *P*=.004) and clinical significance in the intervention group at all time points. The effect size at 12-month follow-up was small to moderate (*d*=0.48). Physical activity increased significantly in the intervention group posttreatment compared with treatment-as-usual group (*P*<.001). In the per-protocol analyses, all secondary measures improved to a greater extent in the intervention group than in the treatment-as-usual group (Table 2).

Participants scoring ≥5 on the PHQ-9 showed greater improvements in all outcome measures in favor of the

intervention group in the intention-to-treat analysis and also in the per-protocol analysis (except for cardiac anxiety at posttreatment) (Table 3). Effect sizes were moderate for CAQ (d=0.55) and EQ-VAS (d=0.71) scores at the 12-month follow-up.

Table 3. Comparison of outcome measures for participants with baseline depression scores (PHQ-9 ≥5).

Outcome	Score, mean (SD)		Intention-to-treat analysis		Per-protocol analysis	
	Intervention[a]	Treatment as usual[b]	95% CI	*P* value[c]	95% CI	*P* value[d]
Primary						
Cardiac Anxiety Questionnaire						
Baseline	25.6 (9.9)	25.6 (10.3)	Reference	N/A[e]	Reference	N/A
Post	20.2 (9.6)	20.7 (10.0)	–0.6 (–3.6 to 2.3)	.66	–0.6 (–3.8 to 2.6)	.71
3 months	16.3 (8.4)	21.8 (10.7)	–3.9 (–6.9 to –0.9)	.01	–4.6 (–7.8 to –1.4)	.006
12 months	15.3 (9.4)	21.3 (12.4)	–4.7 (–7.9 to –1.6)	.003	–5.0 (–8.3 to –1.8)	.002
Body Sensations Questionnaire						
Baseline	39.7 (12.0)	38.6 (14.2)	Reference	N/A	Reference	N/A
Post	32.2 (10.3)	34.0 (13.0)	–3.7 (–7.4 to –0.1)	.04	–4.1 (–8.1 to –0.1)	.046
3 months	30.3 (10.9)	33.0 (12.6)	–4.0 (–7.8 to –0.3)	.04	–5.7 (–9.6 to –1.7)	.006
12 months	31.4 (11.4)	33.1 (14.0)	–3.5 (–7.4 to 0.4)	.08	–4.1 (–8.0 to –0.1)	.04
Secondary						
Patient Health Questionnaire–9						
Baseline	9.4 (4.1)	9.3 (4.4)	Reference	N/A	Reference	N/A
Post	6.4 (4.1)	9.1 (5.0)	–2.4 (–4.0 to –0.8)	.003	–3.1 (–4.8 to –1.4)	.001
3 months	5.6 (3.8)	7.8 (4.4)	–1.7 (–3.3 to –0.0)	.047	–2.3 (–4.1 to –0.6)	.008
12 months	5.9 (4.1)	7.7 (5.3)	–1.3 (–3.0 to 0.4)	.14	–1.8 (–3.5 to –0.1)	.04
EuroQol Visual Analog Scale						
Baseline	59.9 (21.9)	58.3 (20.5)	Reference	N/A	Reference	N/A
Post	65.9 (21.0)	55.2 (21.0)	9.9 (2.1 to 17.7)	.01	14.4 (5.7 to 23.1)	.001
3 months	68.0 (20.4)	54.3 (20.9)	11.3 (3.2 to 19.4)	.007	13.5 (4.7 to 22.2)	.003
12 months	70.9 (21.0)	54.6 (24.8)	12.4 (4.1 to 20.8)	.004	14.2 (5.4 to 22.9)	.002
Physical activity[f]						
Baseline	2.8 (1.7)	3.2 (2.3)	Reference	N/A	Reference	N/A
Post	4.5 (2.0)	3.4 (2.2)	1.3 (0.7 to 2.0)	<.001	1.7 (1.1 to 2.4)	<.001
3 months	3.7 (2.0)	3.4 (2.3)	0.4 (–0.3 to 1.1)	.24	0.6 (–0.1 to 1.2)	.09
12 months	3.7 (2.1)	3.4 (2.3)	0.2 (–0.4 to 0.9)	.48	0.4 (–0.2 to 1.1)	.20

[a]Intervention: baseline (n=51); post (n=40); 3 months (n=35); 12 months (n=34).

[b]Treatment as usual: baseline (n=53); post (n=43); 3 months (n=41); 12 months (n=35).

[c]Estimated mean differences between groups adjusted for baseline values (linear mixed models).

[d]Estimated mean differences between groups adjusted for baseline values (linear mixed models); n=34 in the intervention group and n=35 in the treatment-as-usual group.

[e]N/A: not applicable.

[f]Number of times with moderate physical activity >30 minutes a week.

Adverse Events and Protocol Deviations

The participants of the intervention group did not report any serious adverse events in the telephone contacts with the therapist. Two participants screened had serious depression (scores >20 on the PHQ-9): one patient in the treatment-as-usual group required an acute psychiatric evaluation and was referred to the psychiatric specialist team. The other participant (in the intervention group) wanted to continue with the intervention. In agreement with the participant, the general practitioner was informed of the intervention, and the participant continued the treatment.

Discussion

Principal Findings

In this randomized controlled trial, we compared the effectiveness of internet-based cognitive behavioral therapy and treatment as usual in patients with noncardiac chest pain. Our findings provide strong evidence for the effectiveness of internet-based cognitive behavioral therapy after a negative cardiac evaluation in a hospital setting. Participants receiving internet-based cognitive behavioral therapy experienced improvements in cardiac anxiety, fear of bodily sensations, depressive symptoms, health-related quality of life, and physical activity level. Participants with depressive symptoms at baseline showed greater improvements after internet-based cognitive behavioral therapy on almost all outcome measures compared with those who were not depressed at baseline.

Most (163/231, 70%) patients who were invited agreed to participate, and 84% of participants (68/81) in the intervention group completed at least 5 of the 6 internet-based cognitive behavioral therapy sessions. This indicates that the intervention had a high acceptability. The cumulative therapist time spent for each internet-based cognitive behavioral therapy participant totaled 60 to 70 minutes, including the time spent providing information about the intervention at the hospital. This makes the intervention considerably less costly and time-consuming (per patient) than even brief face-to-face cognitive behavioral therapy.

Primary Outcomes

Cardiac anxiety did not significantly improve at the posttreatment assessment in the internet-based cognitive behavioral therapy group compared with the treatment-as-usual group. However, although the improvements within the intervention group during the intervention period were large, improvements in the treatment-as-usual group were also substantial. This may indicate that an examination by a cardiologist, which is often combined with computed tomography angiography, can contribute to reducing cardiac anxiety. One might expect that improvement in cardiac anxiety in the treatment-as-usual group would be temporary, but the improvements were similarly maintained in the treatment-as-usual group at the 3- and 12-month follow-ups. However, cardiac anxiety in the intervention group improved significantly more than in the treatment-as-usual group at the 3- and 12-month follow-ups in both the intention-to-treat and per-protocol analyses, suggesting that the exclusion of cardiac conditions combined with internet-based cognitive behavioral therapy yield a better effect on cardiac anxiety over time.

As with cardiac anxiety, the improvements in fear of bodily sensations within the intervention group were large and comparable to findings in a previous study [7]. However, the treatment-as-usual group also improved substantially; at the 12-month follow-up, the intervention group reported a nonsignificant improvement of 2.7 points (P=.07) in the intention-to-treat analysis and a statistically significant improvement of 3.1 points (P=.04) in the per-protocol analysis. These improvements in the treatment-as-usual group were somewhat surprising, given that in [7], the treatment-as-usual group actually showed increased BSQ scores at the 12-month follow-up. In [7], the intervention group had substantially and significantly less symptoms on BSQ at 12 months (between-group difference: 7.5 points). The smaller improvement in BSQ in our study might be explained by the fact that the participants in [7] had significant noncardiac chest pain-related complaints at the time of study inclusion (6 months after receiving a noncardiac chest pain diagnosis). Therefore, in [7], the time period most likely associated with a naturalistic remission of noncardiac chest pain symptoms was completed, thus eliminating potential participants with very mild symptoms and little room for improvement from the analysis of key outcome measures.

Secondary Outcomes

PHQ-9 depression scores improved significantly more in the intervention group than in the treatment-as-usual group posttreatment (P=.003) and at the 3-month follow-up (P=.03), but there was no significant difference at the 12-month follow-up (P=.35). It is important to note that the sample had a low baseline score of 6.9 on the PHQ-9, indicating only mild depressive symptoms on average. These low baseline scores reduced the likelihood of finding a significant difference in depression between the internet-based cognitive behavioral therapy and treatment-as-usual groups. Health-related quality of life improved substantially in the intervention group, whereas a slight deterioration was observed in the treatment-as-usual group. This may suggest that even if patients with noncardiac chest pain experience improvement in cardiac anxiety and fear of bodily sensations without specific treatment, they might struggle to regain a normal quality of life. This corresponds well with the findings of other studies [5,25,26]. The improvement in EQ-VAS in the intervention group was statistically significant (posttreatment: P=.009; 3 months: P=.001; 12 months: P=.004) and clinically relevant at all points of time.

There was a significant increase in physical activity in the intervention group posttreatment (P<.001), suggesting that the participants adhered to the physical activity–related elements of our intervention. This initial increase in physical activity in the internet-based cognitive behavioral therapy group was not maintained at the 3- and 12-month follow-up. Clinical impressions suggest that a key element in the internet-based cognitive behavioral therapy was exposure to increased heart rates, followed by the observation that the participants did not experience a health catastrophe. We did not include a specific measure to monitor the intensity of physical activities, but participants who were more physically active also exposed themselves more often to activities that increased their heart rate. We hypothesize that even a temporary increase in activity level with a focus on achieving a high heart rate may have lasting positive effects on cardiac-related anxiety and psychological limitations associated with physical activity.

Subgroup Analysis of Patients With a PHQ-9 Score >5 at Baseline

Studies that included noncardiac chest pain patients regardless of baseline noncardiac chest pain–related symptom levels, such as [8] and our study, might have difficulty detecting significant

differences when baseline scores are low. We found larger effects on most outcome measures among internet-based cognitive behavioral therapy participants with baseline PHQ-9 scores ≥5. Patients with noncardiac chest pain and with depressive symptoms might benefit more from internet-based cognitive behavioral therapy, and PHQ-9 scores should be further analyzed as a possible predictor of treatment effects. Given that internet-based cognitive behavioral therapy was effective for the participant group as a whole, offering a low-cost intervention with no known side effects to all patients with noncardiac chest pain regardless of symptom levels at a busy cardiac unit might be advantageous from an implementation perspective, even though internet-based cognitive behavioral therapy might not be needed for those with low depression scores at baseline.

Comparison With Prior Work

Previously, only one small pilot randomized controlled trial [9] tested the effectiveness of internet-based cognitive behavioral therapy for noncardiac chest pain and did not find statistically significant differences in CAQ or BSQ scores; however, the study was clearly underpowered. In another recent study [8], 3- or 4-session face-to-face cognitive behavioral therapy were compared with treatment as usual in an unselected group of 424 noncardiac chest pain patients recruited at the hospital immediately after their cardiac examination, and health anxiety improved at the 3-month follow-up in the intervention group compared with in the treatment-as-usual group, but these improvements were not sustained at the 12-month follow-up. The main difference between [8] and our study is that our intervention focused on exposure to physical activity (which was not emphasized in [8]). Notably, Jonsbu et al [7] found sustained improvements in BSQ, avoidance of physical activity and depression 12 months after the participants completed their exercise-focused cognitive behavioral therapy treatment. Our focus on physical activity may account for the enduring effects of our intervention. Cognitive behavioral therapy is often more powerful when maladaptive cognitions are challenged during systematic exposure [27].

Strengths and Limitations

This study has several significant strengths. First, a high percentage of potential candidates accepted and participated (162/231, 70%), perhaps because the patients were recruited in a cardiac setting. Second, adherence to the internet-based cognitive behavioral therapy was high (68/81, 84%), and sample retention at the 12-month follow-up was acceptable (117/162, 72.2%). Third, the study was performed in a routine clinical setting with few exclusion criteria, which strengthens the generalizability of the results. Fourth, the intervention required approximately 1 hour of therapist time per patient and was easy to administer. Finally, outcome measures were collected electronically, and randomization was performed electronically at a remote location.

This trial also had some limitations. First, the therapist (TT) who conducted all the check-in calls during this study participated in the development of the intervention and had previous cognitive behavioral therapy training, which may have increased the effect of internet-based cognitive behavioral therapy in this study. In addition, all check-in calls were made by the same therapist, except during a period of a few weeks. Second, the control group only received treatment as usual (we did not use a sham intervention as a control condition). Third, neither the therapist nor the participants could be blinded to the intervention. Fourth, computer literacy and internet access were required in order to utilize the intervention. Finally, the assessment of physical activity and analysis of PHQ-9 ≥5 were not described in the original trial registration. A recent study [8] found that brief cognitive behavioral therapy without a particular focus on physical activity does not have an effect on noncardiac chest pain, suggesting that physical activity is an important aspect of the treatment. The analysis of PHQ-9 ≥5 was added because previous research found that higher depression scores predicted poorer outcomes in patients with noncardiac chest pain [2,28].

Clinical Implications and Further Research

An easily accessible, low-cost, and effective internet-based cognitive behavioral therapy for noncardiac chest pain has the potential to improve care and quality of life in a large and underserved group of patients. Further research is needed to investigate whether standard staff working at the cardiac unit can provide this type of care effectively and to determine whether certain patient subgroups benefit more or less from this type of intervention.

Conclusions

This study provides evidence that internet-based cognitive behavioral therapy with minimal therapist contact and with a focus on physical activity is effective in reducing cardiac anxiety and increasing health-related quality of life in patients with noncardiac chest pain.

Acknowledgments

Southeastern Norway Regional Health Authority (grant 2016031) and Sørlandet Hospital HF supported this project.

We thank Are Hugo Pripp, Gunvor Launes, Ingrid Klovning, and Finn Tore Gjestvang for significant contributions to this project. We also thank the patients who participated in the study and the staff at the cardiac department at Sørlandet Hospital, Kristiansand.

Authors' Contributions

JAH conceived the project. TT, JAH, EWM, LTW, FT, and EJ contributed to study design and intervention development. TT, JAH, EWM, LTW, FT, and EJ contributed to data collection. JAH, FG, and EJ contributed to statistical analysis. JAH, EWM, and EJ prepared the manuscript.

Conflicts of Interest

The technical platform used to deliver the intervention is owned by Assistert Selvhjelp AS. None of the authors has economic relationship with Assistert Selvhjelp AS.

Multimedia Appendix 1
Snapshots from the intervention.
[DOCX File , 3498 KB - jmir_v24i1e33631_app1.docx]

Multimedia Appendix 2
CONSORT-EHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1270 KB - jmir_v24i1e33631_app2.pdf]

References

1. Stallone F, Twerenbold R, Wildi K, Reichlin T, Rubini Gimenez M, Haaf P, et al. Prevalence, characteristics and outcome of non-cardiac chest pain and elevated copeptin levels. Heart 2014 Nov;100(21):1708-1714. [doi: 10.1136/heartjnl-2014-305583] [Medline: 24968886]
2. Jonsbu E, Dammen T, Morken G, Martinsen EW. Patients with noncardiac chest pain and benign palpitations referred for cardiac outpatient investigation: a 6-month follow-up. Gen Hosp Psychiatry 2010;32(4):406-412. [doi: 10.1016/j.genhosppsych.2010.03.003] [Medline: 20633745]
3. Mourad G, Alwin J, Strömberg A, Jaarsma T. Societal costs of non-cardiac chest pain compared with ischemic heart disease--a longitudinal study. BMC Health Serv Res 2013;13:403 [FREE Full text] [doi: 10.1186/1472-6963-13-403] [Medline: 24107009]
4. Mourad G, Strömberg A, Johansson P, Jaarsma T. Depressive symptoms, cardiac anxiety, and fear of body sensations in patients with non-cardiac chest pain, and their relation to healthcare-seeking behavior: a cross-sectional study. Patient 2016 Feb;9(1):69-77. [doi: 10.1007/s40271-015-0125-0] [Medline: 25840677]
5. Mourad G, Alwin J, Jaarsma T, Strömberg A, Johansson P. The associations between psychological distress and health-related quality of life in patients with non-cardiac chest pain. Health Qual Life Outcomes 2020 Mar 12;18(1):68 [FREE Full text] [doi: 10.1186/s12955-020-01297-0] [Medline: 32160887]
6. Kisely SR, Campbell LA, Yelland MJ, Paydar A. Psychological interventions for symptomatic management of non-specific chest pain in patients with normal coronary anatomy. Cochrane Database Syst Rev 2015 Jun 30(6):CD004101 [FREE Full text] [doi: 10.1002/14651858.CD004101.pub5] [Medline: 26123045]
7. Jonsbu E, Dammen T, Morken G, Moum T, Martinsen EW. Short-term cognitive behavioral therapy for non-cardiac chest pain and benign palpitations: a randomized controlled trial. J Psychosom Res 2011 Feb;70(2):117-123. [doi: 10.1016/j.jpsychores.2010.09.013] [Medline: 21262413]
8. Mulder R, Zarifeh J, Boden J, Lacey C, Tyrer P, Tyrer H, et al. An RCT of brief cognitive therapy versus treatment as usual in patients with non-cardiac chest pain. Int J Cardiol 2019 Aug 15;289:6-11. [doi: 10.1016/j.ijcard.2019.01.067] [Medline: 30770262]
9. Mourad G, Strömberg A, Jonsbu E, Gustafsson M, Johansson P, Jaarsma T. Guided internet-delivered cognitive behavioural therapy in patients with non-cardiac chest pain - a pilot randomized controlled study. Trials 2016 Jul 26;17(1):352 [FREE Full text] [doi: 10.1186/s13063-016-1491-1] [Medline: 27456689]
10. Thesen T, Jonsbu E, Martinsen EW, Himle JA, Thorup F, Launes G, et al. Internet-assisted cognitive behavioural therapy for non-cardiac chest pain: a pilot and feasibility study. Cogn Behav Therap 2021 Jul 01;14:1-18. [doi: 10.1017/s1754470x21000155]
11. Eifert GH, Thompson RN, Zvolensky MJ, Edwards K, Frazer NL, Haddad JW, et al. The cardiac anxiety questionnaire: development and preliminary validity. Behav Res Ther 2000 Oct;38(10):1039-1053. [Medline: 11004742]
12. Marker CD, Carmin CN, Ownby RL. Cardiac anxiety in people with and without coronary atherosclerosis. Depress Anxiety 2008;25(10):824-831 [FREE Full text] [doi: 10.1002/da.20348] [Medline: 17597101]
13. Khawaja N. Revisiting the factor structure of the agoraphobic cognitions questionnaire and body sensations questionnaire: a confirmatory factor analysis study. J Psychopathol Behav Assess 2003;25(1):57-63. [doi: 10.1023/A:1022204104267]
14. Van Beek MHCT, Zuidersma M, Lappenschaar M, Pop G, Roest AM, Van Balkom AJLM, et al. Prognostic association of cardiac anxiety with new cardiac events and mortality following myocardial infarction. Br J Psychiatry 2016 Nov;209(5):400-406. [doi: 10.1192/bjp.bp.115.174870] [Medline: 27539297]
15. Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of fear of fear in agoraphobics: the body sensations questionnaire and the agoraphobic cognitions questionnaire. J Consult Clin Psychol 1984 Dec;52(6):1090-1097. [Medline: 6520279]
16. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [Medline: 11556941]

17. Erbe D, Eichert H, Rietz C, Ebert D. Interformat reliability of the patient health questionnaire: Validation of the computerized version of the PHQ-9. Internet Interv 2016 Sep;5:1-4. [doi: 10.1016/j.invent.2016.06.006] [Medline: 25904163]
18. Herdman M, Gudex C, Lloyd A, Janssen M, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). Qual Life Res 2011 Dec;20(10):1727-1736 [FREE Full text] [doi: 10.1007/s11136-011-9903-x] [Medline: 21479777]
19. Kounali D, Button KS, Lewis G, Gilbody S, Kessler D, Araya R, et al. How much change is enough? Evidence from a longitudinal study on depression in UK primary care. Psychol Med 2020 Nov 03:1-8. [doi: 10.1017/S0033291720003700] [Medline: 33138872]
20. Kroenke K, Stump T, Chen C, Kean J, Bair M, Damush TM, et al. Minimally important differences and severity thresholds are estimated for the PROMIS depression scales from three randomized clinical trials. J Affect Disord 2020 Apr 01;266:100-108 [FREE Full text] [doi: 10.1016/j.jad.2020.01.101] [Medline: 32056864]
21. Turkoz I, Alphs L, Singh J, Jamieson C, Daly E, Shawi M, et al. Clinically meaningful changes on depressive symptom measures and patient-reported outcomes in patients with treatment-resistant depression. Acta Psychiatr Scand 2021 Mar;143(3):253-263 [FREE Full text] [doi: 10.1111/acps.13260] [Medline: 33249552]
22. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. Health Qual Life Outcomes 2007 Dec 21;5:70 [FREE Full text] [doi: 10.1186/1477-7525-5-70] [Medline: 18154669]
23. Nolan CM, Longworth L, Lord J, Canavan JL, Jones SE, Kon SSC, et al. The EQ-5D-5L health status questionnaire in COPD: validity, responsiveness and minimum important difference. Thorax 2016 Dec;71(6):493-500 [FREE Full text] [doi: 10.1136/thoraxjnl-2015-207782] [Medline: 27030578]
24. Hu X, Jing M, Zhang M, Yang P, Yan X. Responsiveness and minimal clinically important difference of the EQ-5D-5L in cervical intraepithelial neoplasia: a longitudinal study. Health Qual Life Outcomes 2020 Oct 02;18(1):324 [FREE Full text] [doi: 10.1186/s12955-020-01578-8] [Medline: 33008423]
25. Frieling T. Non-cardiac chest pain. Visc Med 2018 Apr;34(2):92-96 [FREE Full text] [doi: 10.1159/000486440] [Medline: 29888236]
26. Dammen T, Bringager CB, Arnesen H, Ekeberg O, Friis S. A 1-year follow-up study of chest-pain patients with and without panic disorder. Gen Hosp Psychiatry 2006;28(6):516-524. [doi: 10.1016/j.genhosppsych.2006.08.005] [Medline: 17088168]
27. Bilet T, Olsen T, Andersen JR, Martinsen EW. Cognitive behavioral group therapy for panic disorder in a general clinical setting: a prospective cohort study with 12 to 31-years follow-up. BMC Psychiatry 2020 May 24;20(1):259 [FREE Full text] [doi: 10.1186/s12888-020-02679-w] [Medline: 32448206]
28. Mourad G, Jaarsma T, Hallert C, Strömberg A. Depressive symptoms and healthcare utilization in patients with noncardiac chest pain compared to patients with ischemic heart disease. Heart Lung 2012 Sep;41(5):446-455. [doi: 10.1016/j.hrtlng.2012.04.002] [Medline: 22652167]

Abbreviations

BSQ: Body Sensations Questionnaire
CAQ: Cardiac Anxiety Questionnaire
EQ-VAS: EuroQol Visual Analog Scale
PHQ-9: Patient Health Questionnaire–9

Edited by G Eysenbach; submitted 30.09.21; peer-reviewed by E Børøsund; comments to author 19.11.21; revised version received 27.11.21; accepted 30.11.21; published 24.01.22.

Please cite as:
Thesen T, Himle JA, Martinsen EW, Walseth LT, Thorup F, Gallefoss F, Jonsbu E
Effectiveness of Internet-Based Cognitive Behavioral Therapy With Telephone Support for Noncardiac Chest Pain: Randomized Controlled Trial
J Med Internet Res 2022;24(1):e33631
URL: https://www.jmir.org/2022/1/e33631
doi:10.2196/33631
PMID:35072641

Original Paper

The Effectiveness of Sequentially Delivered Web-Based Interventions on Promoting Physical Activity and Fruit-Vegetable Consumption Among Chinese College Students: Mixed Methods Study

Yanping Duan[1,2*], PhD; Wei Liang[1,2*], PhD; Yanping Wang[3], MEd; Sonia Lippke[4], PhD; Zhihua Lin[5], PhD; Borui Shang[6], PhD; Julien Steven Baker[1,2], PhD, DSc

[1]Department of Sport, Physical Education and Health, Faculty of Social Sciences, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[2]Center for Health and Exercise Science Research, Hong Kong Baptist University, Hong Kong, China (Hong Kong)

[3]The National Physical Fitness Lab, Hubei Institute of Sport Science, Wuhan, China

[4]Department of Psychology & Methods, Jacobs University Bremen, Bremen, Germany

[5]Sport Section, Wuhan University, Wuhan, China

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

[*]these authors contributed equally

Corresponding Author:
Yanping Duan, PhD
Department of Sport, Physical Education and Health
Faculty of Social Sciences
Hong Kong Baptist University
12/F Hong Kong Baptist University Shek Mun Campus
8 On Muk Street, Shek Mun, Shatin
Hong Kong, HKG
China (Hong Kong)
Phone: 852 34113038
Fax: 852 34113357
Email: duanyp@hkbu.edu.hk

Abstract

Background: Web-based interventions for multiple health behavior change (MHBC) appear to be a promising approach to change unhealthy habits. Limited research has tested this assumption in promoting physical activity (PA) and fruit-vegetable consumption (FVC) among Chinese college students. Moreover, the timing of MHBC intervention delivery and the order of components need to be addressed.

Objective: This study aims to examine the effectiveness of 2 sequentially delivered 8-week web-based interventions on physical activity, FVC, and health-related outcomes (BMI, depression, and quality of life) and the differences in the intervention effects between the 2 sequential delivery patterns. The study also aims to explore participants' experiences of participating in the health program.

Methods: We conducted a randomized controlled trial, in which 552 eligible college students (mean 19.99, SD 1.04 years, 322/552, 58.3% female) were randomly assigned to 1 of 3 groups: PA-first group (4 weeks of PA followed by 4 weeks of FVC intervention), FVC-first group (4 weeks of FVC followed by 4 weeks of PA intervention), and a control group (8 weeks of placebo treatment unrelated to PA and FVC). The treatment content of two intervention groups was designed based on the Health Action Process Approach (HAPA) framework. A total of four web-based assessments were conducted: at baseline (T1, n=565), after 4 weeks (T2, after the first behavior intervention, n=486), after 8 weeks (T3, after the second behavior intervention, n=420), and after 12 weeks (T4, 1-month postintervention follow-up, n=348). In addition, after the completion of the entire 8-week intervention, 18 participants (mean 19.56, SD 1.04 years, 10/18, 56% female) who completed the whole program were immediately invited to attend one-to-one and face-to-face semistructured interviews. The entire study was conducted during the fall semester of 2017.

Results: The quantitative data supported superior effects on physical activity, FVC, and BMI in the 2 sequential intervention groups compared with the control group. There were no significant differences in physical activity, FVC, and health-related

outcomes between the 2 intervention groups after 8 weeks. The FVC-first group contributed to more maintenance of FVC compared with the PA-first group after 12 weeks. Four major themes with several subthemes were identified in the qualitative thematic analysis: PA and FVC behavior, health-related outcomes, correlates of behavior change, and contamination detection.

Conclusions: This study provides empirical evidence for the effectiveness of sequentially delivered, web-based MHBC interventions on PA and FVC among Chinese college students. The timing issue of MHBC intervention delivery was preliminarily addressed. Qualitative findings provide an in-depth understanding and supplement the quantitative findings. Overall, this study may contribute considerably to future web-based MHBC interventions.

Trial Registration: ClinicalTrials.gov NCT03627949; https://clinicaltrials.gov/ct2/show/NCT03627949

International Registered Report Identifier (IRRID): RR2-10.1186/s12889-019-7438-1

(J Med Internet Res 2022;24(1):e30566) doi:10.2196/30566

KEYWORDS

web-based intervention; physical activity; fruit-vegetable consumption; college students; health action process approach; mixed methods; quantitative research; qualitative research

Introduction

Background

Considerable evidence indicates a high prevalence of physical inactivity and insufficient consumption of fruit and vegetables among college students [1]. In China, especially, recent studies have revealed that about 40% of Chinese college students do not meet the 150 accumulated minutes of moderate physical activity (PA) per week recommended by the World Health Organization [2,3], and more than half of this population does not consume the recommended 5 servings (400 g) of fruit and vegetables per day [4,5].

College students, who are in a crucial transition stage from late adolescence to adulthood, adopting such unhealthy behaviors can increase the risk of many chronic diseases (eg, cardiovascular diseases, obesity, or type 2 diabetes) and jeopardize their mental health (eg, increase the risk of depression) [6,7]. Therefore, promoting PA and fruit-vegetable consumption (FVC) in this population is essential. Over the past 2 decades, much evidence has shown that multiple health behavior change (MHBC) interventions can promote both PA and healthy diet among college students [1,8,9]. Thus, MHBC interventions for college students are promising for supporting long-lasting behavior changes into late adulthood.

With the burgeoning use of the internet, web-based MHBC intervention programs have been widely applied in various populations [10-12]. Compared with traditional face-to-face, analogous-delivered modes, using the internet to deliver MHBC interventions has been demonstrated to have a series of advantages, such as accessibility, scalability, cost-effectiveness, and convenience [13]. For college students, who form the majority of the internet users, the acceptability and effectiveness of web-based MHBC interventions for promoting PA and FVC have been proven by a growing body of evidence [1,9,14]. However, most of the existing studies have been conducted in Western countries, while there is limited research on Chinese college students.

One debatable question in MHBC research is how to deliver MHBC interventions (ie, delivery timing) to achieve more robust treatment effects and whether the order of the sequential intervention contents makes a difference [15]. Opinions differ, and the evidence is limited and inconsistent to date. One view suggests that multiple health behaviors typically coexist as behavioral clusters or bundles [16-19]. For example, one risk behavior (eg, sedentary behavior) often occurs with other risk behaviors (eg, excessive intake of fat and sugar, smoking, or alcohol addiction) or one health-protective behavior (eg, physical activity) coexists with other health-protective behaviors (eg, FVC). This interconnection of health behaviors can generate synergistic or additive effects, contributing to the reinforcement of the treatment effects when changing them simultaneously [16-18]. Furthermore, this approach is shorter and less costly [18]. In contrast, some researchers argue that simultaneous intervention delivery may be overburdened, as it requires individuals to make considerable efforts to self-regulate to adopt multiple health behaviors [20,21]. In addition, a simultaneous approach may not address any particular behavior in sufficient depth, decreasing the potential effects of the intervention. Therefore, sequential delivery may be more suitable [19,20]. This design often requires delivering interventions over a longer period, which potentially increases costs, and if the design lacks motivation strategies, participation and adherence will suffer [22]. In addition to these viewpoints, a review of 6 studies suggested that both simultaneous and sequential approaches can be considered equally effective in MHBC interventions [23]; thus, the main question relates more to the sequential order of the intervention components.

However, previous MHBC studies focused mostly on the combination of heterogeneous categories of behaviors (eg, risk behaviors plus health protecting behaviors). When targeting PA and FVC, which are both health-protecting behaviors, some evidence has indicated that there is a gateway or carry-over effect between PA and FVC [21,24,25]. In a sequential design, changing the first behavior (PA/FVC) may result in increased self-efficacy or self-regulation to undertake the second behavior (FVC/PA), thereby contributing to successful changes in both behaviors [21,24,25]. In addition, using a sequential design may ensure that each behavior can be adequately addressed, and this is particularly important for college students who are at a critical life transition to establish or sustain favorable lifestyles in the long term [8]. Given these findings, it is possible that a sequential approach may be more effective for interventions

that address PA and FVC behavior among college students. Nevertheless, the question related to 2 sequential delivery patterns (PA-first vs FVC-first), of which one would contribute more to improvements in health behaviors and health-related outcomes, is still unclear. Further examination of this aspect is warranted.

This Study

The health action process approach (HAPA) model was used as the theoretical framework of intervention in this study [12,26,27]. The HAPA model postulates two distinctive stages (motivational and volitional) for the process of behavior change, where individuals may experience a dynamic process from generating a behavioral intention to performing and maintaining a specific health behavior [28]. A series of psychosocial factors are considered crucial in the behavioral change process. During the motivational stage, the primary task is to form a behavioral intention, where action self-efficacy, outcome expectancies, and risk perceptions are proposed as contributory antecedents of intention. Once individuals have initiated behavioral intention, volitional factors (action planning and coping planning, maintenance, and recovery self-efficacies) and external resources (social support) play imperative roles in bridging the intention-behavior gap to facilitate behavioral execution and maintenance. The applicability of the HAPA model in promoting various health behaviors among adults has been widely approved [16,29,30].

The mixed methods approach can combine the merits of both quantitative and qualitative methods (ie, integrating the power of stories and the power of numbers) and compensate for their respective limitations [31]. Therefore, this study uses a sequential mixed methods design, including a randomized controlled trial (RCT) and face-to-face interviews. In particular, the RCT aims to quantitatively examine the effectiveness of web-based interventions in promoting PA and FVC behavior and health-related outcomes among Chinese college students. We hypothesize that (1) both intervention groups (PA-first and FVC-first) would show more changes in PA (metabolic equivalent [MET]-min/week) and FVC (Portion/ day) compared with the control group; (2) both intervention groups would show more changes in health-related outcomes (BMI, depression, and perceived quality of life) compared with the control group; and (3) 2 intervention groups would differ significantly in PA (MET-min/week), FVC (portion/day), and health-related outcomes at 8 and 12 weeks. With regard to the one-on-one and face-to-face qualitative interviews, we aim to explore college students' experiences and perceptions of participating in the web-based MHBC intervention program.

Methods

Quantitative Study: An RCT

Participants and Procedure

Considering the feasibility and limited resources, an RCT was adopted instead of a cluster RCT with a standard 3-arm, parallel, double-blinded design. The study participants were undergraduate students from a university in the central region of China. The sample size was calculated using G*Power software (Version 3.1). As this study aimed to improve MHBC, PA and FVC were treated as *coprimary outcomes*, and the family-wise error rate was not necessary to be controlled for the sample size calculation [32,33]. Considering the effect size of PA change was smaller than that of FVC in our previous studies, the sample size estimate in this study was based on an average effect size of 0.45 on PA only (Cohen *d*) [26,29] to ensure robust analyses for the treatment effects on both outcomes (ie, conjunctive power) [32,33]. Therefore, to provide a power of 0.8 (1-β) with an α of .05, 79 participants were required for each group. Assuming a dropout rate of 40% to the posttest (dropout rates ranged from approximately 30% to 50% in previous studies [26]), a total of 396 participants (132 per group) were needed.

A total of 634 college students were contacted in their first general physical education (PE) classes with the assistance of PE lecturers during the fall semester of 2017. In China, college students are required to take PE courses based on the national education guideline [34]. Therefore, it was feasible to recruit students from different departments via their PE classes. Once students expressed interest, they were provided with a hard copy of the study consent form and were invited to complete the web-based registration (including sociodemographic information) by scanning a quick response (QR) code. Among the 634 participants, 565 (89.1%) college students (aged ≥18 years) met the following eligibility criteria: (1) were not collegiate athletes or majoring in any sport-related subjects, (2) were not vegetarians, (3) had no restrictions on physical mobility (eg, heart diseases, stroke, or disability) or FVC (eg, fruit allergies or diabetes), and (4) were able to use a computer or laptop and mobile phone and had access to the internet.

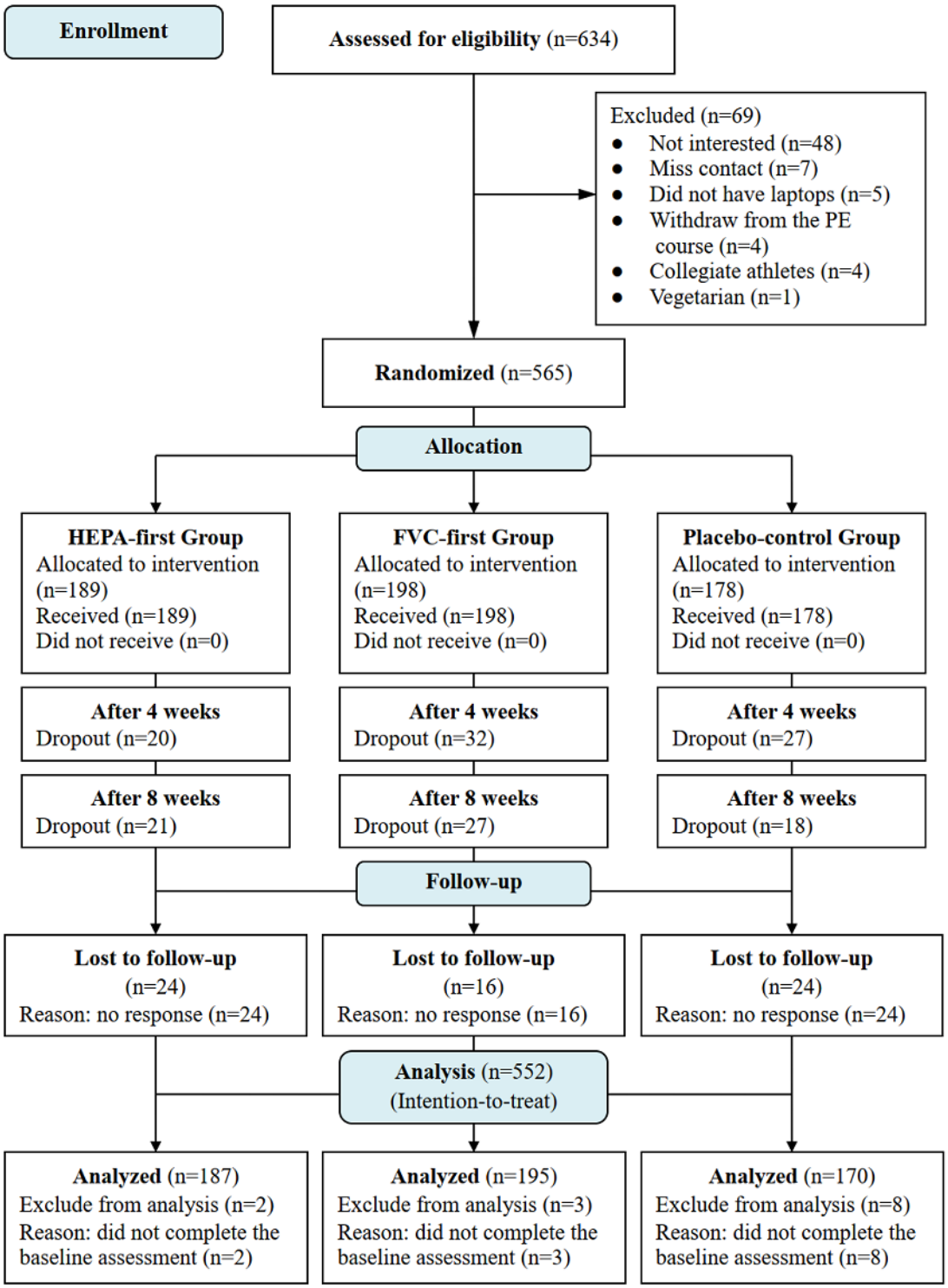
After qualification screening, all 565 eligible students were randomly allocated to 1 of 3 groups. The random number list was generated at the backend management system of the website platform. The three groups consisted of intervention group 1 (PA-first group: first 4-week intervention addressing physical activity, followed by a 4-week intervention addressing FVC; 189/565, 33.5%), intervention group 2 (FVC-first group: first 4-week intervention addressing FVC, followed by a 4-week intervention addressing physical activity; 198/565, 35.0%), and a placebo control group (8-week placebo treatments, which were not relevant to changing actual PA and FVC behavior; 178/565, 31.5%). One week after randomization, participants were invited to attend the session once a week for approximately 25 minutes each time for 8 weeks. All participants were asked to complete the weekly health sessions independently and not to discuss the content of the health sessions with their classmates, roommates, and friends.

In addition to attending the intervention session, all participants were asked to complete electronic questionnaires at 4 time points: baseline (T1, at the beginning of the intervention), after 4 weeks (T2, after the completion of the 4-week intervention on the first behavior), after 8 weeks (T3, after the completion of the 4-week intervention on the second behavior), and after 12 weeks (T4, 1-month follow-up after intervention completion; Multimedia Appendix 1). After excluding participants who did not complete the baseline assessment, the final sample consisted

of 552 participants. Participant flow and retention are shown in Figure 1.

We obtained ethical approval from the research ethics committee of Hong Kong Baptist University (FRG2/15-16/032) and registered the study on ClinicalTrials.gov (NCT03627949).

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram. FVC: fruit-vegetable consumption; PA: physical activity; PE: physical education.



Intervention

The intervention was designed based on the health action process approach (HAPA) model, which lasted for 8 weeks [9,12,27,28]. The PA-first module was designed in a previous project conducted by our research team [26,29]. In particular, the intervention content of the first 4-week period targeted the following HAPA-based constructs for physical activity: week 1 included risk perception, outcome expectancies, and goal settings (contributing to the formation and enhancement of PA

intention); week 2 included development of action planning; week 3 included revision and adjustment of previous action planning and development of coping planning; and week 4 included revision and adjustment of previous coping planning and development of perceived social support. The same intervention constructs were then implemented to target the HAPA-based social cognitive determinants of FVC change in the second 4-week period. The earlier project only addressed one sequence of intervention delivery (PA-first) among Chinese college students [24]. This study extended the intervention delivery modules by adding an FVC-first module. Multimedia Appendix 2 shows the intervention variables and content for the 2 intervention groups and a placebo control group. Furthermore, a series of behavior change techniques were used in the weekly intervention sessions to facilitate the implementation and maintenance of PA and FVC [35] (Multimedia Appendix 3).

In addition, given the high dropout rate (31.6% for posttest and 71.2% for the follow-up test) in our previous study [24], several new strategies and approaches were implemented in this study to increase the attractiveness of the health program and maximize participant retention [36], including (1) restructuring the website page to match the preferences of young adults, for example, supplementing vivid pictures and redrawing the layout; (2) adding pop-up messages to prevent participants from missing a unit or a single item; and (3) using WeChat (a popular social media platform in China) to contact the participants and deliver the reminder messages. In addition, participants who adhered to the 8-week intervention and 4 time-point data collection would be provided with monetary incentive (US $8.0) as a reward.

Measures

Primary Outcomes: PA and FVC

PA was measured using the Chinese short version of the International Physical Activity Questionnaire (IPAQ-C) [37]. The IPAQ-C asked participants to report the frequency and duration of engaging in 3 intensities of PA in the past 7 days. The weekly MET-min of total PA were calculated using the following equation [37]:

PA (MET-min/week) = 8.0 METs × vigorous intensity of PA (min/week) + 4.0METs × moderate intensity of PA (min/week) + 3.3METs × light intensity of PA (min/week)

FVC was measured by using a 4-item scale, asking the participants to count the portions of fruit and vegetables they consumed on average during a typical day [38]. Each item (*raw vegetables*, *fruit*, *cooked or steamed vegetables*, and *fruit or vegetable juice*) had 11 options for the number of portions such as 0, 0.5, 1, 1.5, 2, and 2.5 until 5 or above. The total consumed portion was the sum of each item.

Secondary Outcomes: Health-Related Outcomes

BMI was measured by asking participants to self-report their body weight (kg) and height (m). BMI was calculated using the formula *BMI = weight (kg)/height squared (m^2)* [39].

Depression was measured using the Chinese version of the Centre for Epidemiologic Studies Depression–10 Scale [40]. Participants were asked with the question stem as “In the past week how often I feel” followed by 10 items such as “...I was bothered by things that usually don’t bother me” (Cronbach α=.78). Answers were given on a visual analog scale (VAS) from rarely or none of the time (<1 day) *0* to most or all the time (5-7 days) *3* [40,41].

Perceived quality of life was assessed using the short version of the World Health Organization Quality of Life-BREF [42]. Respondents were first asked about their general quality of life as “How would you rate your quality of life?” with a VAS score ranging from (very poor=1) to (very good=5). The physical health subdomain with 7 items was also measured (Cronbach α=.71), such as “How satisfied are you with your ability to perform your daily living activities?” with a VAS score ranging from (very dissatisfied=1) to (very satisfied=5).

Demographic Information

Gender, age, relationship status (single or in a relationship), grade (freshman or sophomore or junior or senior), major, and self-reported health status (bad or medium or good) were included in demographic information.

All measurement instruments were translated into a Chinese version and validated in previous studies with Chinese college students [26]. Demographic information was collected only at registration (T0). All other measurements were measured at T1, T2, T3, and T4.

Statistical Analyses

The data were analyzed using SPSS (version 25.0; IBM Corporation). Analyses of variances, independent *t* tests, and chi-square tests were performed to examine whether the randomization was successful. Where there were significant differences across groups at baseline, the variables were treated as covariates in subsequent analyses. Data analyses for the intervention effects were performed using the intention-to-treat principle, with per-protocol analysis as a sensitivity test [43]. Missing values were addressed using the multiple imputation method with chained equations, except for dropouts, which were imputed with the baseline-observation-carried-forward approach [44]. A series of generalized linear mixed models were applied to evaluate the intervention effects on outcome measures at different time points, with time, group, and their interaction as fixed effects, and with individuals as random effects. On the basis of the −2 log likelihood and Akaike and Bayesian information criteria, an unstructured covariance structure was selected for the model estimation using a restricted maximum likelihood approach. For the post hoc comparison, considering that studies with co-primary outcomes may increase the type-II error rate and decrease the study power [33], the least significant difference method was used rather than other adjusted approaches (eg, Bonferroni) [45-47]. In addition, chi-square tests were used for post hoc comparisons. The 5% level (2-tailed) was used as the statistical significance cut-off point.

Qualitative Study: Interviews

Study Design and Participants

On the basis of the guideline of descriptive phenomenology, a series of one-on-one and face-to-face semistructured interviews were conducted, which involved three types of questions (open-ended, closed-ended, and conformational) [48].

To reach theoretical saturation, based on the *rule of thumb* and *calculating the mean of selected qualitative studies*, 18 participants who completed the web-based MHBC interventions were randomly recruited from 2 intervention groups and 1 control group in the RCT (6 participants for each group).

Procedure and Data Collection

The research team jointly developed an interview guide (4 experts in the health psychology domain), including questions, prompts, and guides, based on suggestions from Bryman and Flick [48,49]. A detailed interview guide can be found elsewhere [50] (Multimedia Appendix 4).

One-on-one and face-to-face semistructured interviews were conducted after the completion of the 8-week web-based MHBC interventions. On the basis of the interview guide, the main question was used to invite the participants to talk freely, such as "What is your experience with the 8-week web-based health program?" Additional questions were asked during the conversation for clarification and elaboration, such as "If so, can you explain in more detail? What caused this change in your behavior?" Each interview was audio-recorded and lasted approximately 30 minutes. Each interviewee could obtain US $6.5 as a participation remuneration if they completed the interview.

Data Analysis

The audio-recorded interview data were transcribed orthographically and organized using NVivo (version 11; QSR International). Thematic analysis was used for data analysis, including six phases: familiarization with the data, generation of the initial codes, searching for themes, reviewing the potential themes, defining and naming themes, and producing the report [51]. Two members of the research team (WL and YW) independently conducted the first five steps (intercoder agreement=97%). All discrepancies in any aspect of the analysis process (eg, defining the potential themes and subthemes) were discussed by three members of the research team (WL, YW, and YD) until consensus was reached. Both inductive and deductive processes were involved in the thematic analysis process. Finally, to guarantee the credibility and trustworthiness of the qualitative study, the entire procedure followed a set of principles, including sensitivity, commitment, rigor, transparency, coherence impact, and importance (Multimedia Appendix 5) [52]. To guarantee that this study complied with qualitative reporting standards, the 32-item Consolidated Criteria for Reporting Qualitative Studies (COREQ) checklist was used [53] (Multimedia Appendix 6).

Results

Quantitative Study Results

Randomization Check and Sample Characteristics

A randomization check indicated that there were no significant differences in baseline characteristics across the 3 groups in relation to gender, age, study year, relationship status, and self-reported health status (P=.37-.83). Moreover, the 3 groups did not differ significantly in all continuous and categorical variables with regard to the physical activity, FVC, and health outcomes at baseline (P=.098-.93). Therefore, the randomization was successful.

A description of the study sample is provided in Table 1. A total of 552 valid respondents were recruited from 28 different departments (the total number of university departments is 34), including 322 (58.3%) females and 230 (41.7%) males, with the age ranging from 18 to 24 years (mean 19.99, SD 1.04 years). Most of the participants were freshmen and sophomores, that is 47.8% (264/552) and 41.5% (229/552) of the total sample, respectively. Among these participants, only 46 (8.3%) indicated being in a relationship. The average BMI was 20.41 kg/m^2 (SD 2.45 kg/m^2). Most participants (374/552, 67.8%) had a healthy weight, 23% (127/552) were underweight, and 9.2% (51/552) were overweight. In addition, 64.9% (358/552) of the participants indicated a medium level of self-reported health status.

Table 1. Sociodemographic information, PA[a], FVC[b], and health outcomes of the study sample at baseline.

Variable	Total (N=552)	PA-first (n=187)	FVC-first (n=195)	Control (n=170)
Sociodemographic information				
Age (range 18-24 years), mean (SD)	19.99 (1.04)	20.07 (1.07)	19.96 (0.99)	19.93 (1.06)
Gender, n (%)				
Male	230 (41.7)	79 (42.2)	78 (40)	73 (42.9)
Female	322 (58.3)	108 (57.8)	117 (60)	97 (57.1)
Grade, n (%)				
Freshman	264 (47.8)	86 (46.0)	90 (46.2)	88 (51.8)
Sophomore	229 (41.5)	77 (41.2)	84 (43.1)	68 (40)
Junior	46 (8.3)	18 (9.6)	16 (8.2)	12 (7.1)
Senior	13 (2.4)	6 (3.2)	5 (2.6)	2 (1.2)
Marital status, n (%)				
Single	506 (91.7)	170 (90.9)	183 (93.8)	153 (90)
In a relationship	46 (8.3)	17 (9.1)	12 (6.2)	17 (10)
Health status, n (%)				
Poor	17 (3)	5 (2.7)	9 (4.6)	3 (1.8)
Medium	358 (64.9)	122 (65.2)	125 (64.1)	111 (65.3)
Good	177 (32.1)	60 (32.1)	61 (31.3)	56 (32.9)
PA and FVC , mean (SD)				
PA (MET[c]-min/week)	2124.23 (1244.42)	2180.22 (1314.89)	2074.45 (1191.03)	2119.73 (1229.34)
FVC (Portion/day)	3.81 (1.75)	3.84 (1.70)	3.82 (1.87)	3.76 (1.68)
Health-related outcomes				
BMI (range 15.62-32.88 kg/m^2), mean (SD)	20.41 (2.45)	20.32 (2.34)	20.52 (2.62)	20.40 (2.39)
Depression, mean (SD)	0.92 (0.69)	0.85 (0.63)	0.93 (0.73)	0.98 (0.72)
Quality of life, mean (SD)	3.15 (0.67)	3.23 (0.63)	3.14 (0.67)	3.08 (0.71)

[a]PA: physical activity.

[b]FVC: fruit-vegetable consumption.

[c]MET: metabolic equivalent.

Intervention Effects on PA and FVC

Table 2 presents the results of the evaluation of the intervention effects on the weekly amount of PA and daily servings of FVC. Figure 2A and Figure 2B show the descriptive information of the 2 behaviors from T1 to T4. The results revealed that both health behaviors changed favorably and significantly over time (all $P \leq .001$) and that there were significant differences in the time and treatment effects between the intervention and control groups in both PA ($P=.02$) and FVC ($P<.001$) behaviors. From the post hoc tests, we found small effect sizes for the intervention effects on PA (Cohen $d=0.22$-0.29) and small-to-medium effect sizes on FVC behavior (Cohen $d=0.34$-0.59).

To identify which intervention delivery schedule would be more effective in promoting health behavior change after 8 and 12 weeks, we compared the differences in each health behavior between the PA-first and FVC-first groups at T3 and T4. The results indicated that the 2 intervention groups did not differ significantly from each other in PA at either time point, but there was a significant difference in FVC between the 2 intervention groups ($P=.014$). The group receiving FVC instruction first had a significantly higher FVC after 12 weeks (T4).

Table 2. Results of the generalized linear mixed models with physical activity and fruit–vegetable consumption after 4, 8, and 12 weeks as outcome measures (n=552).

Time and group	PA[a]			FVC[b]		
	Values	*P* value	Effect size, Cohen *d*	Values	*P* value	Effect size, Cohen *d*
Type III tests, F_{549}						
Time×group	2.66	.015	N/A[c]	12.17	<.001	N/A
Time	5.24	.001	N/A	36.40	<.001	N/A
Group	2.05	.13	N/A	13.64	<.001	N/A
After 4 weeks (T2)[d], mean difference						
PA-first vs control	231.58	.041	0.22	0.19	.47	0.08
FVC-first vs control	109.70	.33	0.10	1.42	<.001	0.58
PA-first vs FVC-first	121.87	.26	0.11	−1.23	<.001	0.51
After 8 weeks (T3)[d], mean difference						
PA-first vs control	282.36	.018	0.25	1.13	<.001	0.44
FVC-first vs control	321.19	.007	0.29	1.35	<.001	0.52
PA-first vs FVC-first	−38.83	.74	0.03	−0.22	.41	0.08
After 12 weeks (T4)[d], mean difference						
PA-first vs control	253.21	.026	0.24	0.81	.02	0.34
FVC-first vs control	252.39	.025	0.24	1.41	<.001	0.59
PA-first vs FVC-first	0.83	.99	<0.01	−0.60	.014	0.25

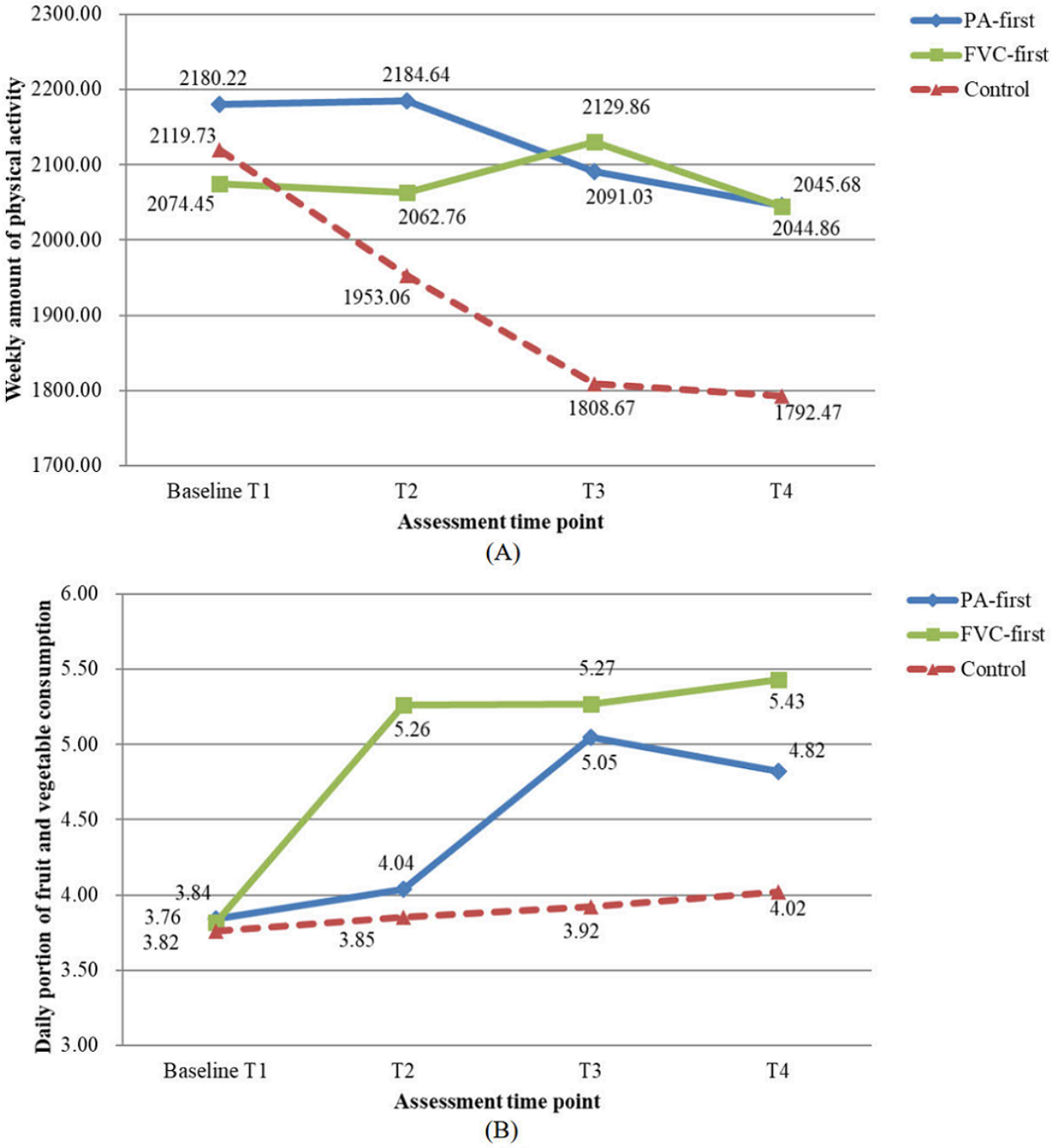
[a]PA: physical activity (metabolic equivalent of task-min/week).

[b]FVC: fruit-vegetable consumption (portion/day).

[c]N/A: not applicable.

[d]Post hoc test: least significant difference; mean difference was significant at the .05 level.

Figure 2. Mean values for 3 groups from timepoints T1 to T4. (A) weekly amount of physical activity (metabolic equivalent of task-min/week). (B) daily portion of fruit and vegetable consumption (portion/day). FVC: fruit-vegetable consumption; PA: physical activity.



Intervention Effects on Health-Related Outcomes

The descriptive results revealed that the intervention groups had a favorable time effect on health-related outcomes compared with the control condition (all $P<.01$; Table 3). For the time and treatment interaction, the difference was only significant for BMI ($P=.03$), and there were no significant differences between the intervention and control groups for depression ($P=.60$) and perceived quality of life ($P=.07$). From the post hoc tests, we found small effect sizes for the intervention effects on BMI (Cohen $d=0.01$-0.07), depression (Cohen $d=0.07$-0.31), and quality of life (Cohen $d=0.07$-0.47). Descriptive information is presented in Figure 3A-Figure 3C.

Figure 3. Mean values for 3 groups from timepoints T1 to T4. (A) BMI (kg/m2). (B) Depression. (C) Quality of life. FVC: fruit-vegetable consumption; PA: physical activity.

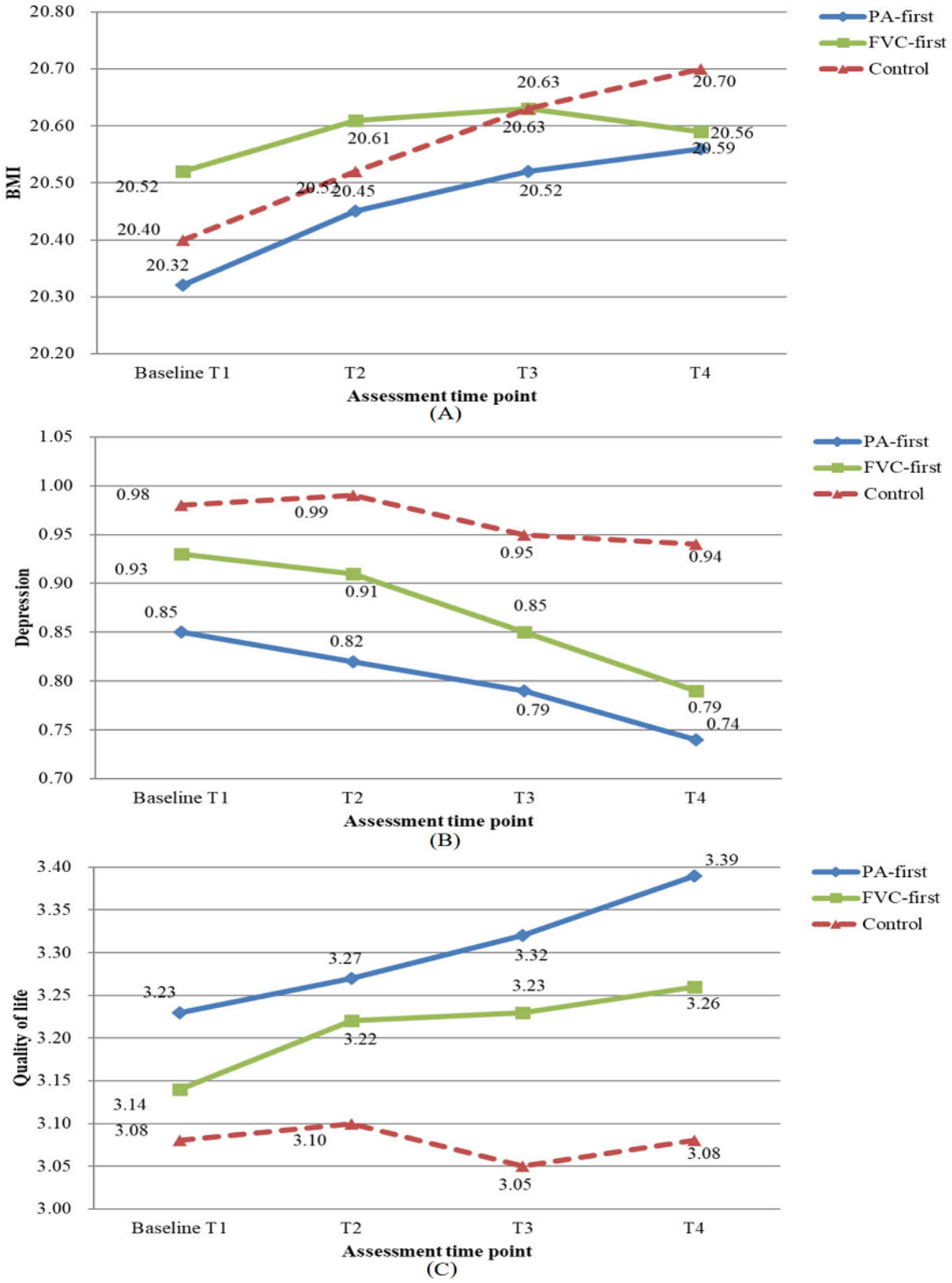


Table 3. Results of the generalized linear mixed models with health-related outcomes (ie, BMI, depression, and perceived quality of life) after 4, 8, and 12 weeks as outcome measures (n=552).

Time and group	BMI			Depression			Quality of life		
	Value	*P* value	Effect size, Cohen *d*	Value	*P* value	Effect size, Cohen *d*	Value	*P* value	Effect size, Cohen *d*
Type III tests, F_{549}									
Time×group	2.34	.03	N/A[a]	.76	.60	N/A	1.95	.07	N/A
Time	18.29	<.001	N/A	8.21	<.001	N/A	6.69	<.001	N/A
Group	.15	.86	N/A	3.61	.03	N/A	6.69	.001	N/A
After 4 weeks (T2)[b], mean difference									
PA[c]-first vs control	−0.08	.76	0.03	−0.17	.01	0.27	0.16	.01	0.27
FVC[d]-first vs control	0.09	.71	0.04	−0.07	.28	0.11	0.12	.07	0.18
PA-first vs FVC-first	−0.17	.49	0.07	−0.10	.14	0.17	0.05	.46	0.07
After 8 weeks (T3)[b], mean difference									
PA-first vs control	−0.11	.64	0.05	−0.16	.02	0.25	0.27	<.001	0.42
FVC-first vs control	−0.02	.92	0.01	−0.10	.13	0.15	0.18	.009	0.26
PA-first vs FVC-first	−0.09	.70	0.04	−0.06	.37	0.10	0.09	.18	0.15
After 12 weeks (T4)[b], mean difference									
PA-first vs control	−0.14	.56	0.06	−0.19	.003	0.31	0.30	<.001	0.47
FVC-first vs control	−0.10	.66	0.04	−0.15	.02	0.23	0.18	.01	0.26
PA-first vs FVC-first	−0.04	.88	0.01	−0.05	.45	0.07	0.12	.07	0.19

[a]N/A: not applicable.

[b]Post hoc test: least significant difference; mean difference is significant at the .05 level.

[c]PA: physical activity (metabolic equivalent of task-min/week).

[d]FVC: fruit-vegetable consumption (portion/day).

Dropout Analysis and Sensitivity Test

The dropout rate of participants was 14.3% (79/552) from T1 to T2, 14% (66/473) from T2 to T3, and 15.7% (64/407) from T3 to T4. The aggregated dropout percentage was 26.3% (145/552) from T1 to T3, and 37.9% (209/552) from T1 to T4. There were no between-group differences in the percentage of participants with incomplete data at T2, T3, and T4 (χ^2_2=1.3-3.4, *P*=.18-.51).

Overall, sensitivity analyses exhibited similar results for all dependent variables, except for the perceived quality of life, and the time and treatment effect was found to be statistically significant (*P*=.03) for both intervention groups compared with the control group (Multimedia Appendix 7). This indicated that our intervention significantly improved the perceived quality of life only for participants who adhered to the entire intervention. However, as noncompliance is inevitable in the real world, the results of the sensitivity test did not alter the findings of the primary analyses of the nonsignificant effect of time and treatment on the perceived quality of life (this was also consistent with the results of the likelihood-based estimation).

Qualitative Study Results

Sample Characteristics

A total of 18 participants attended the interview study, including 10 females and 8 males, ranging in age from 18 to 22 years (mean 19.56, SD 1.04 years); 89% (16/18) of the participants were single; 61% (11/18) of the participants reported a medium level of health status, while 39% (7/18) indicated a good level of health status. Table 4 presents the demographic characteristics of the participants.

Table 4. Demographic information of interviewees (n=18).

Participant ID	Group	Gender	Age (years)	Marital status	Health status
1	IG1[a]	Female	19	Single	Medium
2	IG1	Male	20	Single	Medium
3	IG1	Male	19	Single	Medium
4	IG1	Male	19	Single	Medium
5	IG1	Female	20	Single	Good
6	IG1	Female	19	Single	Medium
7	IG2[b]	Male	18	Single	Good
8	IG2	Male	19	Single	Medium
9	IG2	Male	18	Single	Good
10	IG2	Female	20	Single	Medium
11	IG2	Female	19	Single	Medium
12	IG2	Female	21	In a relationship	Medium
13	CG[c]	Female	21	Single	Good
14	CG	Female	20	Single	Good
15	CG	Female	19	Single	Medium
16	CG	Male	20	Single	Good
17	CG	Female	22	Single	Good
18	CG	Male	19	In a relationship	Medium

[a]IG1: physical activity–first group.

[b]IG2: food-vegetable consumption–first group.

[c]CG: control group.

Major Themes

All participants were invited to talk about their experiences and participation in the web-based MHBC intervention program. Through thematic analysis, four major themes were identified: (1) PA and FVC behavior, (2) health-related outcomes, (3) correlates of health behavior change, and (4) contamination detection.

Theme 1: PA and FVC Behavior

This theme focused on how the students self-assessed their current status and changes in PA and FVC behavior over the previous 8 weeks. It contained three subthemes: (1) improving health behaviors, (2) no change in health behaviors, and (3) decrease in health behaviors.

Improving Health Behaviors

In total, 4 of the 12 (33%) students in the intervention groups reported improvement in their PA in the last 8 weeks (participants 1, 2, 4, and 6), while no student in the control group indicated an increase in this behavior. In addition, 2 students with improved PA stated that they understood the importance of performing sufficient PA through the web-based health program (participants 2 and 6). They noted some improvements, even after encountering obstacles at the start of implementation. For example:

> *I think the health program is quite helpful...I had not paid much attention to the health issue, especially since the last year of high school, you know, I was busy with my studies...But recently, thanks to the health learning sessions, I started to worry about my health status and I really improved this behavior these days, I feel that doing some physical activity makes all of my days...* [Participant 6]

For FVC, the feedback was more positive in the intervention groups, as 10 of the 12 (83%) students (except participants 1-4) described their improvement in the consumption of fruit and vegetables per day after receiving the health interventions. In comparison, only one student (participant 18) in the control group reported an increase in this behavior:

> *I eat more fruit and vegetables every day after participating in the health learning program...I pay more attention to this health issue now.* [Participant 9]

No Change in Health Behaviors

The results revealed that 5 of the 12 (41%) students (participants 3, 5, 7, 8, and 10) in the intervention groups reported maintaining their physical activity, while 2 of the 6 (33%) students in the control group reported no change in their weekly amount of PA (participants 17 and 18). The participants explained the following:

> *In the last two months, there has been no prominent change in this behavior...I maintain the same intensity and the same amount of weekly physical activity.* [Participant 3]

For FVC, two students (participants 1 and 4) in the intervention groups reported no change in their daily consumption of fruit and vegetables compared with four students in the control condition group (participants 13, 14, 16, and 17).

Decrease in Health Behaviors

Of the 12 students, 3 (25%) students in the intervention groups reported reducing the weekly amount of PA for diverse reasons, such as weather and study-related activities (participants 9, 11, and 12). Of the 6, 4 (66%) students in the control group also witnessed a decrease in PA in the last 8 weeks (participants 13-16). For FVC, only one student in the control group reported reducing her daily consumption of fruit and vegetables (participant 15): "I used to eat apples or grapes every day, but now I do not because of the freezing weather and other reasons."

Theme 2: Health-Related Outcomes

This theme focused on how students self-assessed their physical and mental health outcomes. It contained three subthemes: *body weight*, *depression*, and *perceived quality of life*.

Body Weight

Of the 12 students, 4 (33%) in the intervention groups showed an increase in body weight (participants 2, 4, 9, and 11), and 3 of the 6 students (50%) in the control group described a similar trend (participants 13, 15, and 16). Two students in the intervention groups explained that their body weight increased because they were fitter and had more muscles these days (participants 2 and 4), but no further elaboration and explanation was obtained from the students in the control group. For example:

> *My body weight has increased a bit, but I think it is due to my muscles...there is no obvious change in my body fat.* [Participant 2]

Depression

Most of the students experienced no symptoms of depression in the previous 8 weeks (12/18, 66%). Although no student described the change in their level of depression, most students recognized the positive effects of PA and FVC in reducing depression. In particular, six students mentioned that PA can help fight depression (participants 2, 3, 5, 10, 17, and 18). Two students (participants 2 and 12) described the positive effect of FVC on reducing depression and felt that this influence was weaker than that of physical activity. For instance:

> *Doing exercise is useful for coping with depression and eating fruit and vegetables can put me in a good mood...But, I think that dietary behavior is not as effective as exercise in dealing with this problem...* [Participant 12]

Quality of Life

In total, 10 of the 18 (55%) students felt that their quality of life was good before participating in the health program. A total of 10 respondents (4 in the intervention groups and 6 in the control group) indicated no change in their quality of life. Two respondents indicated a decrease in this aspect (participants 9 and 11), and 6 respondents in the intervention groups (participants 2, 3, 5, 6, 8, and 12) felt that they were more energetic and their perceived well-being improved after participating in the web-based health program. In total, 9 of the 18 (50%) students recognized that consuming enough fruit and vegetables could help, while 6 students indicated the positive influence of regular PA on improving their perceived quality of life (Multimedia Appendix 8). For instance:

> *After participating in the health learning program, I exercised more, it brought me a good spiritual outlook...I felt that I slept better...Greasy food made me uncomfortable and fresh fruit and vegetables improved my well-being.* [Participant 2]

Theme 3: Correlates of Behavior Change

This theme reflected the students' narratives of the correlates of health behavior change. It contained two subthemes: university policy for PA and barriers to PA and FVC.

University Policy for Physical Activity

The respondents mentioned that to encourage students to engage in PA, their university had a policy in place, named *Ham Run*, which was used as one of the assignments in the physical education class. The university declared that all undergraduates had to run 2000 m 28 times, accounting for 20% of the course credit. The students also indicated that they had to complete the *Ham Run* task before November, as the physical education course final usually takes place in mid-November:

> *The Ham Run task was okay for me, I completed it at the end of October...I feel that I was less active after completing this task.* [Participant 5]

Barriers to PA and FVC

*The s*tudents reported several barriers to their motivation and implementation of health behaviors. For physical activity, students mentioned the weather and facilities that inhibited their engagement in physical activity. For example:

> *The weather is freezing and the playground is quite far from my dormitory, so I rarely go outside to exercise...* [Participant 9]

> *Sometimes I want to exercise indoors, but the venues are always unavailable...it is really difficult.* [Participant 3]

For FVC, the barriers included the weather, the supply of university canteens, and financial issues. For instance, some students explained the following:

> *I do not want to eat vegetables because the weather is freezing and my teeth need hot food.* [Participant 15]

> *I cannot choose what I want to eat...it is decided by the university canteens...I hope they can improve their supply of vegetables.* [Participant 4]

> *The fruit sold nearby is quite expensive...I can only afford one serving of fruit per day...I do not want to ask my family for more money.* [Participant 1]

Theme 4: Contamination Detection

This theme consisted of three subthemes related to potential intervention contamination: *communication with classmates in the same physical education class*; *communication with classmates in a different physical education class*; and *communication with friends, roommates, and family*.

Communication With Classmates in the Same Physical Education Class

All students in the intervention and control groups indicated that they did not discuss the content of the health program with other classmates in the same PE class. The students provided additional information, emphasizing that there was a low possibility for students from the same department to enroll in the same PE class to discuss the content due to the university's curriculum selection system. Some students explained the following:

> *I did not discuss the content with others, as I am not familiar with my classmates and we come from different departments...it is impossible for us (students in my department or roommates) to select the same PE class...* [Participant 5]

Communication With Students in a Different Physical Education Class

All students explained that they did not discuss the content with other students participating in the health program but were enrolled in a different PE class. For instance:

> *I do not know if the students enrolled in other PE classes were also invited to join the health program...I will not communicate with others, even if I find some acquaintances participating in this program.* [Participant 7]

Communication With Friends, Roommates, and Family

In total, 17 of the 18 students (94%) reported that they did not discuss the health program with their friends, roommates, and family. Only 1 student told her parents about her participation in the health program, but did not discuss the content of the intervention:

> *I told my mother that I participated in a health learning program and she encouraged me to adhere to it...but I did not give details.* [Participant 2]

Discussion

Principal Findings

For the quantitative part of the study, most of the research hypotheses were supported. For the qualitative analysis, 4 main themes with a couple of subthemes were identified through thematic analysis. The qualitative findings corresponded to the quantitative findings, providing an in-depth understanding of changes in PA and FVC behavior in Chinese college students.

Intervention Effects on Behavioral Indicators

The principal expected intervention effects on the behavioral indicators of PA and FVC were identified in the quantitative study. From the findings of the RCT, compared with students in the placebo control condition, students in both intervention groups reported significant and favorable changes in the weekly amount of PA and daily consumption of fruit and vegetables, which supported hypothesis 1. The findings were more positive than those of our previous study, in which a significant treatment effect was only supported for FVC change in Chinese college students [24]. The treatment effects in this study were consistent with another study, which was conducted in Germany and the Netherlands, that aimed to improve the PA and FVC in adults who intended to reduce cardiovascular risk [54]. As the intervention materials were similar to those used in previous studies [54], the intervention effects on PA and FVC found in this study might be suitable for use in other Asian and European countries.

In terms of the differences in intervention effects on behavioral changes between the 2 delivery timings (PA-first vs FVC-first), hypothesis 3 was partially supported in the RCT. We found that there was no significant difference in PA between the 2 intervention groups at either time point, whereas the FVC-first group had significantly higher consumption of daily fruit and vegetables after 12 weeks than the PA-first group. Our finding is partially consistent with a previous study in middle-aged adults, which found that the PA-first group showed higher PA than the diet-first group (Cohen d=0.37, P<.001), while the diet-first group showed a higher FVC than the PA-first group (Cohen d=0.28, P<.001) after a 12-month intervention [23]. Our findings might be interpreted as compensatory effects [55]. It seems that FVC-first students in this study were more likely to consume more fruit and vegetables to compensate for the reduced PA compared with PA-first students. Nevertheless, this assumption has not been systematically examined in this study and deserves further investigation.

Intervention Effects on Health-Related Outcomes

The findings of the RCT revealed that both intervention groups showed more changes in BMI compared with the control group, which partially supported hypothesis 2. As most of the participants were freshmen and sophomores and the data collection (ie, T3 and T4) was conducted near the beginning of the winter holidays, most of the participants increased their body weight because of the seasonal time that the study was conducted (special transition stage of life and seasonal variation) [56]. However, the students in the 2 intervention groups had significant reductions in the upward trend in BMI compared with the students in the control condition. This was consistent with findings from other MHBC studies, which suggested that combined PA and diet interventions had a more robust effect on weight management than interventions on either PA or diet alone in adults [57].

For depression, in the RCT, we did not find a significant effect, probably because of the floor effect [58], that is, the college students in this study reported a low incidence of depression (mean 0.92, SD 0.69; scale range from 0 to 3) at the beginning of the intervention. However, to prevent depression over the long term, more components that include stress management techniques and explicitly address mental health problems need to be developed and examined in future studies. The average score for quality of life was relatively high (3.15, SD 0.67; scale

range from 1 to 5), which validated the baseline assessment of a high level of PA and a low level of depression among the participants. The ceiling effect might lead to the nonsignificance of the intervention on the perceived quality of life, coupled with the depression indicator, indicating the healthy mental states of these participants.

Dropout Rate

It is not surprising that the dropout rates in this study were significantly lower than those in our previous study at both the postintervention test (26.3% vs 31.6%) and at the 1-month follow-up (37.9% vs 71.2%) [24]. As the psychosocial constructs and behavior change techniques were the same in our 2 studies, the higher retention rates might be attributable to the strategies applied in this study (eg, improvement of website design and technology and multiple reminders), which is consistent with other researchers' suggestions [36]. However, the intervention and follow-up durations were comparatively short in this study, and the dropout issue still needs to be considered in future MHBC intervention programs.

Qualitative Findings

The 4 themes provided a clear picture of the participants' experiences and perceptions of participating in the web-based MHBC intervention program. The first theme reflected the perceived changes in PA and FVC of participants. Most of the students in the intervention groups indicated a favorable change in these 2 behaviors, especially for FVC (10/12, 83% indicated an increase). In contrast, most students in the control group showed a decrease in PA (4/6, 66%) and no improvement or decline in FVC (5/6, 83%). Theme 2 reflects participants' perceptions of their change in health-related outcomes. Around half of the students mentioned an upward trend in body weight (7/18, 38%), most of whom indicated a decrease in PA (5/7, 71%). A total of 2 students who improved PA also showed a slight increase in their body weight and attributed it to muscle enhancement. No statistically significant effects on depression and quality of life were identified from the quantitative data, and we found that most students reflected a good knowledge of the benefits of adequate PA and FVC in depression (5/12, 41%) and quality of life (10/12, 83%) after receiving the web-based MHBC interventions. Despite no explicitly positive comments about the change in these 2 indicators, half of the students in the intervention groups described that they became more invigorated and had a better perceived quality of life after participating in the web-based health interventions.

From the qualitative study, we also found some additional information that underlined the crucial role of university policy in promoting physical activity and revealed prominent barriers to PA and FVC behavior. The students described the university's relevant policy, which motivated their engagement in physical activity. To some extent, this can explain the situation in the previous RCT of participants reporting a relatively high amount of weekly physical activity. These results also echoed the suggestions from other studies, emphasizing the importance of including *sports time* in curricula and suggesting that supportive school policies should be considered when promoting the health of college students [5,59]. Furthermore, in the qualitative study, the students highlighted the extrinsic environmental factors obstructing the execution of PA and FVC, such as weather, facilities, and financial support. These barriers are consistent with findings from a previous qualitative study conducted in the United Kingdom, emphasizing the university environment and finance as barriers to students' PA and dietary behavior [59].

Finally, theme 4 reflected the results of contamination detection, which provided qualitative evidence and explanation for a low risk of contamination in this study.

Strength and Limitations

This study has considerable theoretical and practical implications for further web-based MHBC interventions. The use of the mixed methods approach increased the external validity of the quantitative data for factor-outcome relationship and thus was generalizable to a larger college student population and also ensured the strong internal validity of the in-depth descriptive qualitative data regarding complex context-specific issues and phenomena (eg, participating in a web-based MHBC intervention program) [60]. Despite the methodological merits and profound implications of the study findings, several limitations need to be addressed. First, all the research data were obtained via self-report measures or narratives, which may lead to recall bias and social desirability effects [61]. Furthermore, although a physical education course at a university in China provided a convenient setting for the RCT design, spillover and contamination could not be ignored [62]. Several strategies have been used to minimize this problem, and the follow-up interview did not identify any contamination; however, this issue still deserves further consideration (eg, adopting a cluster RCT). Furthermore, the intervention study was conducted during winter, where the effects might be confounded by seasonal factors. Further examination, including different seasonal contexts, is warranted. As with all qualitative research, the findings generated in this study could not be regarded as representative of all student samples who received a web-based MHBC intervention for PA and FVC, and caution is needed when generalizing from the interview sample to wider populations. In addition, given previous evidence and theoretical assumptions, this study focused only on a sequential delivery mode, comparing PA-first with FVC-first. However, to more comprehensively address the timing of MHBC intervention delivery, further studies should add a simultaneous design to compare the advantages of the 2 delivery timings and also consider participants' preferences [63].

Conclusions

Using a mixed methods approach, the study demonstrated the potential of a web-based and theory-based MHBC intervention for promoting both PA and FVC among Chinese college students. Moreover, the differences in intervention effects on changes in PA and FVC between the 2 delivery sequences were primarily identified in the quantitative study. In addition, the qualitative interviews provided an in-depth understanding of the quantitative findings and identified PA policy and external barriers as other determinants of change in PA and FVC. The overall findings provide new insights into MHBC research, providing theoretical and practical implications for future design and the application of web-based MHBC interventions.

Acknowledgments

This research was supported by the Faculty Research Grant of the Hong Kong Baptist University (FRG2/15-16/032). The funding organization had no role in the study design, study implementation, manuscript preparation, or publication decision. This work is the responsibility of the authors [42].

Authors' Contributions

YD, WL, and SL conceived and designed the study. YD, WL, and YW contributed to the preparation of the study materials. YD, WL, and YW collected the data. WL, YW, and YD screened and analyzed the data. WL and YD drafted and revised the manuscript. All authors have reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study process.
[PNG File , 286 KB - jmir_v24i1e30566_app1.png]

Multimedia Appendix 2
Intervention variables and content for the 2 intervention groups and the setting for a placebo control group.
[DOCX File , 24 KB - jmir_v24i1e30566_app2.docx]

Multimedia Appendix 3
Behavioral change techniques.
[DOCX File , 17 KB - jmir_v24i1e30566_app3.docx]

Multimedia Appendix 4
Interview guide.
[DOCX File , 17 KB - jmir_v24i1e30566_app4.docx]

Multimedia Appendix 5
Trustworthiness of the qualitative study.
[DOCX File , 14 KB - jmir_v24i1e30566_app5.docx]

Multimedia Appendix 6
Consolidated Criteria for Reporting Qualitative Studies (COREQ) checklist (32-item).
[DOCX File , 18 KB - jmir_v24i1e30566_app6.docx]

Multimedia Appendix 7
Sensitivity test using a per-protocol strategy.
[DOCX File , 14 KB - jmir_v24i1e30566_app7.docx]

Multimedia Appendix 8
A summary of the qualitative findings.
[DOCX File , 15 KB - jmir_v24i1e30566_app8.docx]

Multimedia Appendix 9
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 658 KB - jmir_v24i1e30566_app9.pdf]

References

1. Oosterveen E, Tzelepis F, Ashton L, Hutchesson MJ. A systematic review of eHealth behavioral interventions targeting smoking, nutrition, alcohol, physical activity and/or obesity for young adults. Prev Med 2017;99:197-206. [doi: 10.1016/j.ypmed.2017.01.009] [Medline: 28130046]

2. Global recommendations on physical activity for health. World Health Organization. 2010. URL: https://www.who.int/dietphysicalactivity/global-PA-recs-2010.pdf [accessed 2022-01-11]
3. Xu X, Pu Y, Sharma M, Rao Y, Cai Y, Zhao Y. Predicting physical activity and healthy nutrition behaviors using social cognitive theory: cross-sectional survey among undergraduate students in Chongqing, China. Int J Environ Res Public Health 2017 Nov 05;14(11):1346 [FREE Full text] [doi: 10.3390/ijerph14111346] [Medline: 29113089]
4. Fruit and vegetables for health : report of the Joint FAO/WHO Workshop on Fruit and Vegetables for Health. World Health Organization. 2004. URL: https://apps.who.int/iris/handle/10665/43143 [accessed 2022-01-11]
5. Li X, Yang H, Yang F. Influences of campus environment on physical activity participation of college students. J Wuhan Inst Phys Educ (in Chinese) 2018;52:74-81. [doi: 10.15930/j.cnki.wtxb.2018.01.012]
6. Aune D, Giovannucci E, Boffetta P, Fadnes LT, Keum N, Norat T, et al. Fruit and vegetable intake and the risk of cardiovascular disease, total cancer and all-cause mortality-a systematic review and dose-response meta-analysis of prospective studies. Int J Epidemiol 2017;46(3):1029-1056 [FREE Full text] [doi: 10.1093/ije/dyw319] [Medline: 28338764]
7. Dietz WH, Douglas CE, Brownson RC. Chronic disease prevention: tobacco avoidance, physical activity, and nutrition for a healthy start. J Am Med Assoc 2016;316(16):1645-1646. [doi: 10.1001/jama.2016.14370] [Medline: 27668419]
8. Plotnikoff RC, Costigan SA, Williams RL, Hutchesson MJ, Kennedy SG, Robards SL, et al. Effectiveness of interventions targeting physical activity, nutrition and healthy weight for university and college students: a systematic review and meta-analysis. Int J Behav Nutr Phys Act 2015;12:45 [FREE Full text] [doi: 10.1186/s12966-015-0203-7] [Medline: 25890337]
9. Liang W, Duan YP, Shang BR, Wang YP, Hu C, Lippke S. A web-based lifestyle intervention program for Chinese college students: study protocol and baseline characteristics of a randomized placebo-controlled trial. BMC Public Health 2019;19(1):1097 [FREE Full text] [doi: 10.1186/s12889-019-7438-1] [Medline: 31409381]
10. Duan Y, Shang B, Liang W, Du G, Yang M, Rhodes RE. Effects of ehealth-based multiple health behavior change interventions on physical activity, healthy diet, and weight in people with noncommunicable diseases: systematic review and meta-analysis. J Med Internet Res 2021;23(2):e23786 [FREE Full text] [doi: 10.2196/23786] [Medline: 33616534]
11. Duff OM, Walsh DM, Furlong BA, O'Connor NE, Moran KA, Woods CB. Behavior change techniques in physical activity eHealth interventions for people with cardiovascular disease: systematic review. J Med Internet Res 2017;19(8):e281 [FREE Full text] [doi: 10.2196/jmir.7782] [Medline: 28768610]
12. Lippke S, Fleig L, Wiedemann AU, Schwarzer R. A computerized lifestyle application to promote multiple health behaviors at the workplace: testing its behavioral and psychological effects. J Med Internet Res 2015;17(10):e225 [FREE Full text] [doi: 10.2196/jmir.4486] [Medline: 26429115]
13. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
14. Epton T, Norman P, Dadzie A, Harris PR, Webb TL, Sheeran P, et al. A theory-based online health behaviour intervention for new university students (U@Uni): results from a randomised controlled trial. BMC Public Health 2014;14:563 [FREE Full text] [doi: 10.1186/1471-2458-14-563] [Medline: 24903620]
15. Geller K, Lippke S, Nigg CR. Future directions of multiple behavior change research. J Behav Med 2017;40(1):194-202. [doi: 10.1007/s10865-016-9809-8] [Medline: 27785652]
16. Lippke S, Nigg CR, Maddock JE. Health-promoting and health-risk behaviors: theory-driven analyses of multiple health behavior change in three international samples. Int J Behav Med 2012;19(1):1-13 [FREE Full text] [doi: 10.1007/s12529-010-9135-4] [Medline: 21234735]
17. Södergren M, McNaughton SA, Salmon J, Ball K, Crawford DA. Associations between fruit and vegetable intake, leisure-time physical activity, sitting time and self-rated health among older adults: cross-sectional data from the WELL study. BMC Public Health 2012;12:551 [FREE Full text] [doi: 10.1186/1471-2458-12-551] [Medline: 22830932]
18. Prochaska JJ, Prochaska JO. A review of multiple health behavior change interventions for primary prevention. Am J Lifestyle Med 2011;5(3):208-221 [FREE Full text] [doi: 10.1177/1559827610391883] [Medline: 24358034]
19. King AC, Castro CM, Buman MP, Hekler EB, Urizar GG, Ahn DK. Behavioral impacts of sequentially versus simultaneously delivered dietary plus physical activity interventions: the CALM trial. Ann Behav Med 2013;46(2):157-168 [FREE Full text] [doi: 10.1007/s12160-013-9501-y] [Medline: 23609341]
20. Persky I, Spring B, Vander Wal JS, Pagoto S, Hedeker D. Adherence across behavioral domains in treatment promoting smoking cessation plus weight control. Health Psychol 2005;24(2):153-160. [doi: 10.1037/0278-6133.24.2.153] [Medline: 15755229]
21. Fleig L, Kerschreiter R, Schwarzer R, Pomp S, Lippke S. 'Sticking to a healthy diet is easier for me when I exercise regularly': cognitive transfer between physical exercise and healthy nutrition. Psychol Health 2014;29(12):1361-1372. [doi: 10.1080/08870446.2014.930146] [Medline: 24894668]
22. Prochaska JJ, Spring B, Nigg CR. Multiple health behavior change research: an introduction and overview. Prev Med 2008;46(3):181-188 [FREE Full text] [doi: 10.1016/j.ypmed.2008.02.001] [Medline: 18319098]

23. James E, Freund M, Booth A, Duncan MJ, Johnson N, Short CE, et al. Comparative efficacy of simultaneous versus sequential multiple health behavior change interventions among adults: a systematic review of randomised trials. Prev Med 2016;89:211-223. [doi: 10.1016/j.ypmed.2016.06.012] [Medline: 27311332]
24. Miao M, Gan Y, Gan T, Zhou G. Carry-over effect between diet and physical activity: the bottom-up and top-down hypotheses of hierarchical self-efficacy. Psychol Health Med 2017;22(3):266-274. [doi: 10.1080/13548506.2016.1160134] [Medline: 26965658]
25. Jayawardene WP, Torabi MR, Lohrmann DK. Exercise in young adulthood with simultaneous and future changes in fruit and vegetable intake. J Am Coll Nutr 2016;35(1):59-67. [doi: 10.1080/07315724.2015.1022268] [Medline: 26251968]
26. Duan YP, Wienert J, Hu C, Si GY, Lippke S. Web-based intervention for physical activity and fruit and vegetable intake among Chinese university students: a randomized controlled trial. J Med Internet Res 2017;19(4):e106 [FREE Full text] [doi: 10.2196/jmir.7152] [Medline: 28396306]
27. Schwarzer R. Modeling health behavior change: how to predict and modify the adoption and maintenance of health behaviors. Applied Psychol 2008;57(1):1-29. [doi: 10.1111/j.1464-0597.2007.00325.x]
28. Schwarzer R. Health Action Process Approach (HAPA) as a theoretical framework to understand behavior change. Actualidades en Psicología 2016;30(121):119-130. [doi: 10.15517/ap.v30i121.23458]
29. Duan YP, Liang W, Guo L, Wienert J, Si GY, Lippke S. Evaluation of a web-based intervention for multiple health behavior changes in patients with coronary heart disease in home-based rehabilitation: pilot randomized controlled trial. J Med Internet Res 2018;20(11):e12052 [FREE Full text] [doi: 10.2196/12052] [Medline: 30455167]
30. Zhang C, Zhang R, Schwarzer R, Hagger MS. A meta-analysis of the health action process approach. Health Psychol 2019;38(7):623-637. [doi: 10.1037/hea0000728] [Medline: 30973747]
31. Ivankova NV, Creswell J, Stick SL. Using mixed-methods sequential explanatory design: from theory to practice. Field Methods 2006;18(1):3-20. [doi: 10.1177/1525822X05282260]
32. Vickerstaff V, Omar R, Ambler G. Methods to adjust for multiple comparisons in the analysis and sample size calculation of randomised controlled trials with multiple primary outcomes. BMC Med Res Methodol 2019;19(1):129 [FREE Full text] [doi: 10.1186/s12874-019-0754-4] [Medline: 31226934]
33. Multiple endpoints in clinical trials guidance for industry. The United States Food and Drug Administration. 2017. URL: https://www.fda.gov/files/drugs/published/Multiple-Endpoints-in-Clinical-Trials-Guidance-for-Industry.pdf [accessed 2022-01-11]
34. Notice on Printing and Distributing the "Outline of Teaching Guidance for Physical Education Courses in National Colleges and Universities". Ministry of Education of the People's Republic of China. URL: http://www.moe.gov.cn/s78/A10/moe_918/tnull_8465.html [accessed 2022-01-11]
35. Michie S, Richardson M, Johnston M, Abraham C, Francis J, Hardeman W, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: building an international consensus for the reporting of behavior change interventions. Ann Behav Med 2013;46(1):81-95. [doi: 10.1007/s12160-013-9486-6] [Medline: 23512568]
36. Ludden GD, van Rompay TJ, Kelders SM, van Gemert-Pijnen JE. How to increase reach and adherence of web-based interventions: a design research viewpoint. J Med Internet Res 2015;17(7):e172 [FREE Full text] [doi: 10.2196/jmir.4201] [Medline: 26163456]
37. Macfarlane DJ, Lee CC, Ho EY, Chan KL, Chan DT. Reliability and validity of the Chinese version of IPAQ (short, last 7 days). J Sci Med Sport 2007;10(1):45-51. [doi: 10.1016/j.jsams.2006.05.003] [Medline: 16807105]
38. Rafferty AP, Anderson JV, McGee HB, Miller CE. A healthy diet indicator: quantifying compliance with the dietary guidelines using the BRFSS. Prev Med 2002;35(1):9-15. [doi: 10.1006/pmed.2002.1056] [Medline: 12079436]
39. WHO Expert Consultation. Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet 2004;363(9403):157-163. [doi: 10.1016/S0140-6736(03)15268-3] [Medline: 14726171]
40. Rankin SH, Galbraith ME, Johnson S. Reliability and validity data for a Chinese translation of the Center for Epidemiological Studies-Depression. Psychol Rep 1993;73(3 Pt 2):1291-1298. [doi: 10.2466/pr0.1993.73.3f.1291] [Medline: 8115582]
41. Liang W, Duan Y, Shang B, Hu C, Baker JS, Lin Z, et al. Precautionary behavior and depression in older adults during the COVID-19 pandemic: an online cross-sectional study in Hubei, China. Int J Environ Res Public Health 2021;18(4):1853 [FREE Full text] [doi: 10.3390/ijerph18041853] [Medline: 33672885]
42. Yao G, Wu C. Similarities and differences among the Taiwan, China, and Hong-Kong versions of the WHOQOL questionnaire. Soc Indic Res 2009;91(1):79-98. [doi: 10.1007/s11205-008-9326-4]
43. Thabane L, Mbuagbaw L, Zhang S, Samaan Z, Marcucci M, Ye C, et al. A tutorial on sensitivity analyses in clinical trials: the what, why, when and how. BMC Med Res Methodol 2013;13:92 [FREE Full text] [doi: 10.1186/1471-2288-13-92] [Medline: 23855337]
44. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons Inc; 2004.
45. Althouse A. Adjust for multiple comparisons? It's not that simple. Ann Thorac Surg 2016;101(5):1644-1645. [doi: 10.1016/j.athoracsur.2015.11.024] [Medline: 27106412]
46. Liu S, Tu D. On the applications of Fisher's Least Significant Difference (LSD) procedure in three-arm clinical trials with survival endpoints. Drug Information J 2008;42(1):81-91 [FREE Full text] [doi: 10.1177/009286150804200112]

47. Lin Y. Commentary on the use of Bonferroni approach in post-hoc adjustion. Taipei Medical University: Biostatisticas Research Center. 2015. URL: http://biostat.tmu.edu.tw/page/epaper/ep_download/5stat.pdf [accessed 2022-01-14]
48. Flick U. An Introduction to Qualitative Research. Thousand Oaks, California: SAGE Publications; 2018.
49. Bryman A. Social Research Methods. Oxford, United Kingdom: Oxford University Press; 2016.
50. Liang W. Web-based Sequentially Delivered Interventions on Health-enhancing Physical Activity and Fruit-vegetable Consumption in Chinese College Students. Kowloon Tong, Hong Kong: Hong Kong Baptist University; 2020.
51. Braun V, Clarke V, Weate P. Using thematic analysis in sport and exercise research. In: Routledge Handbook of Qualitative Research in Sport and Exercise. London, United Kingdom: Routledge; 2016:213-227.
52. Yardley L. Demonstrating the validity of qualitative research. J Positive Psychol 2017;12(3):295-296. [doi: 10.1080/17439760.2016.1262624]
53. Booth A, Hannes K, Harden A, Noyes J, Harris J, Tong A. COREQ (consolidated criteria for reporting qualitative studies). In: Guidelines for Reporting Health Research: A User's Manual. Hoboken, New Jersey: Wiley; 2014:214-226.
54. Storm V, Dörenkämper J, Reinwand DA, Wienert J, De Vries H, Lippke S. Effectiveness of a web-based computer-tailored multiple-lifestyle intervention for people interested in reducing their cardiovascular risk: a randomized controlled trial. J Med Internet Res 2016;18(4):e78 [FREE Full text] [doi: 10.2196/jmir.5147] [Medline: 27068880]
55. Lippke S, Dahmen A, Gao L, Guza E, Nigg CR. To what extent is internet activity predictive of psychological well-being? Psychol Res Behav Manag 2021;14:207-219 [FREE Full text] [doi: 10.2147/PRBM.S274502] [Medline: 33642884]
56. Stok FM, Renner B, Clarys P, Lien N, Lakerveld J, Deliens T. Understanding eating behavior during the transition from adolescence to young adulthood: a literature review and perspective on future research directions. Nutrients 2018;10(6):667 [FREE Full text] [doi: 10.3390/nu10060667] [Medline: 29794986]
57. Johns DJ, Hartmann-Boyce J, Jebb SA, Aveyard P, Behavioural Weight Management Review Group. Diet or exercise interventions vs combined behavioral weight management programs: a systematic review and meta-analysis of direct comparisons. J Acad Nutr Diet 2014;114(10):1557-1568 [FREE Full text] [doi: 10.1016/j.jand.2014.07.005] [Medline: 25257365]
58. Garin O. Ceiling effect. In: Michalos AC, editor. Encyclopedia of Quality of Life and Well-Being Research. Dordrecht, the Netherlands: Springer; 2014:631-633.
59. Deliens T, Deforche B, De Bourdeaudhuij I, Clarys P. Determinants of physical activity and sedentary behaviour in university students: a qualitative study using focus group discussions. BMC Public Health 2015;15:201 [FREE Full text] [doi: 10.1186/s12889-015-1553-4] [Medline: 25881120]
60. Best practices for mixed methods research in the health sciences. Implementation Science e-hub. URL: https://implementationscience-gacd.org/wp-content/uploads/2020/11/Best-Practices-for-Mixed-Methods-Research-in-the-Health-Sciences-2018-01-25-1.pdf [accessed 2022-01-11]
61. Chan DK, Stenling A, Yusainy C, Hikmiah Z, Ivarsson A, Hagger MS, et al. Editor's Choice: consistency tendency and the theory of planned behavior: a randomized controlled crossover trial in a physical activity context. Psychol Health 2020;35(6):665-684. [doi: 10.1080/08870446.2019.1677904] [Medline: 31642350]
62. Nezu AM, Nezu CM, editors. Evidence-Based Outcome Research: A Practical Guide to Conducting Randomized Controlled Trials for Psychosocial Interventions. Oxford, United Kingdom: Oxford University Press; 2007.
63. Schulz DN, Kremers SP, Vandelanotte C, van Adrichem MJ, Schneider F, Candel MJ, et al. Effects of a web-based tailored multiple-lifestyle intervention for adults: a two-year randomized controlled trial comparing sequential and simultaneous delivery modes. J Med Internet Res 2014;16(1):e26 [FREE Full text] [doi: 10.2196/jmir.3094] [Medline: 24472854]

Abbreviations

FVC: fruit-vegetable consumption
HAPA: health action process approach
MHBC: multiple health behavior change
PA: physical activity
PE: physical education
RCT: randomized controlled trial
VAS: visual analogue scale

Edited by G Eysenbach; submitted 20.05.21; peer-reviewed by S Ghanvatkar, F Tzelepis; comments to author 11.06.21; revised version received 21.07.21; accepted 19.12.21; published 26.01.22.

Please cite as:
Duan Y, Liang W, Wang Y, Lippke S, Lin Z, Shang B, Baker JS
The Effectiveness of Sequentially Delivered Web-Based Interventions on Promoting Physical Activity and Fruit-Vegetable Consumption Among Chinese College Students: Mixed Methods Study
J Med Internet Res 2022;24(1):e30566
URL: https://www.jmir.org/2022/1/e30566
doi:10.2196/30566
PMID:35080497

Original Paper

Quality of Life and Multilevel Contact Network Structures Among Healthy Adults in Taiwan: Online Participatory Cohort Study

Tso-Jung Yen[1*], PhD; Ta-Chien Chan[2*], PhD; Yang-Chih Fu[3*], PhD; Jing-Shiang Hwang[1*], PhD

[1]Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

[2]Research Center for Humanities and Social Sciences, Academia Sinica, Taipei, Taiwan

[3]Institute of Sociology, Academia Sinica, Taipei, Taiwan

[*]all authors contributed equally

Corresponding Author:
Tso-Jung Yen, PhD
Institute of Statistical Science
Academia Sinica
Environmental Changes Research Building
No.128, Academia Road, Section 2, Nankang
Taipei, 11529
Taiwan
Phone: 886 0981451477
Email: tjyen@stat.sinica.edu.tw

Abstract

Background: People's quality of life diverges on their demographics, socioeconomic status, and social connections.

Objective: By taking both demographic and socioeconomic features into account, we investigated how quality of life varied on social networks using data from both longitudinal surveys and contact diaries in a year-long (2015-2016) study.

Methods: Our 4-wave, repeated measures of quality of life followed the brief version of the World Health Organization Quality of Life scale (WHOQOL-BREF). In our regression analysis, we integrated these survey measures with key time-varying and multilevel network indices based on contact diaries.

Results: People's quality of life may decrease if their daily contacts contain high proportions of weak ties. In addition, people tend to perceive a better quality of life when their daily contacts are face-to-face or initiated by others or when they contact someone who is in a good mood or someone with whom they can discuss important life issues.

Conclusions: Our findings imply that both functional and structural aspects of the social network play important but different roles in shaping people's quality of life.

(J Med Internet Res 2022;24(1):e23762) doi:10.2196/23762

KEYWORDS

contact diary; egocentric networks; social support; weak ties; World Health Organization Quality of Life Survey; quality of life; networks; demography; society

Introduction

People's quality of life (QoL) diverges not just on where they stand in their life cycle and the socioeconomic hierarchy but also on how they are connected with others. Overall, well-connected persons tend to perceive a better quality of their lives [1-11]. Following various definitions and instruments of social networks, existing studies have verified such positive links between social networks and QoL. For example, the structure of the social network is highly correlated with QoL among female workers with high levels of stress at home and work [12]. It also differs significantly between cancer patients with high QoL and those with low QoL [13]. In addition, QoL of older adults varies significantly by the structure of their social networks in terms of the existence of a spouse, size of their family, contact frequency of family members, and network of friends [14]. Moreover, when patients with mental illness discuss their health issues more often with their social network members, it is more likely that they will recover from the disease [15].

Although these studies make various contributions to the literature, they tend to focus on specific occasions or groups. It

is relatively unknown how social networks are associated with QoL in everyday life or among healthy and young adults. In addition, previous studies on QoL tended to measure people's QoL or social networks based on cross-sectional instruments. When measuring personal networks, some studies are further limited by focusing on network size only. Although network size is an important measure of a social network's global structure, it reveals little about the social network's local structure. In other words, network size is insufficient for reflecting how people interact with others in their social networks and how they feel about one another during social interactions. Other than network structures, such functional aspects of social networks are also important for understanding how the social networks are associated with QoL [5].

In this study, we examined how QoL varies in both structure and function of social networks, by analyzing diary data of healthy young adults' social contacts. We adopted a repeated measurement design and conducted a 4-wave survey to measure participants' QoL using the brief version of the World Health Organization Quality of Life Survey (WHOQOL-BREF). In addition, we tracked participants' daily social interactions using in-depth contact diaries [16-18]. We used these diary data to build archives about comprehensive personal networks [19] that reveal participants' social networking in everyday life. The diary data also allow us to pay attention to the functional aspects of the social networks, especially how people feel about each specific interpersonal contact. By integrating both survey and diary data, we were able to investigate how social networks are associated with QoL at contact, tie, and individual levels. Following this bottom-up approach [20], along with multiple-wave measures and multilevel research designs, we aimed to better capture how social networks shape people's daily life experiences.

Methods

Recruitment and Data Collection

We recruited participants and collected their survey and diary data via an online platform called ClickDiary [16,17]. Based on a web-based design, ClickDiary helps researchers implement and manage diary studies and lowers the burden for research participants to record repetitive and unique content relevant to various contact situations. To facilitate compatible analysis, we extracted 2 data sets collected between November 25, 2015 and November 22, 2016. The first data set consisted of a 4-wave survey of participants' QoL measured using the WHOQOL-BREF [21]. This measurement has been further modified to take local social and cultural norms into account [22,23]. We implemented the 4-wave survey on November 25, 2015; February 23, 2016; August 23, 2016; and November 22, 2016. Figure 1 shows a schematic plot of the data collection procedure. In our analysis, we only focused on healthy participants. We excluded from our data set participants who had at least one of the following chronic diseases: cardiovascular disease, diabetes, and cancer. The reason we excluded these participants is that, although it is well-known that these chronic diseases play important roles in shaping people's QoL, among 154 participants in our original data set, only 5 participants had these chronic diseases. The sample size was not large enough to validate the importance of these chronic diseases on people's QoL. After excluding the 5 participants, all remaining 149 participants in our data set did not have these chronic diseases. Among the participants, 34 completed the survey in all 4 waves, 51 completed 3 waves, 28 completed 2 waves, and 36 completed only 1 of the 4 waves. In total, the participants generated 358 records of the survey. The participants in our sample were mainly young adults with an average age of 36 years. Of 149 participants, 110 (73.8%) were female. Being skewed in terms of age and gender, a typical bias present in in-depth diary studies [19], our sample was not representative by any means, and we thus refrained from making any inference from our findings about survey data. Instead, we focused on how the survey results could benefit from detailed information in the diary data. The statistics on age and gender, along with other personal attributes, are summarized in Table 1.

The second data set consisted of detailed contact records nested at 3 unique, yet intertwined, levels taken from ClickDiary: individual characteristics of the participant (that is, the diary keeper, or "ego" in the participant's egocentric network), unique attributes of the contact person ("alter") or the relationship between ego and alter ("tie"), and the situation of each specific contact between each ego-alter pair ("contact"). Each participant recorded contact information and health information in ClickDiary on a daily basis. To better align such information with the survey measures, we constructed social network attributes using only the contact information the participants recorded 2 weeks before they completed a wave of the WHOQOL-BREF survey. These social network attributes in turn facilitated the analysis of how they could help determine QoL.

Because a typical in-depth contact diary study requires its participants to devote themselves to recording interpersonal contacts in all forms for months, and sometimes over a year, the task of diary-keeping has been so demanding and tedious that only a small group of volunteers could possibly finish the study even with reasonable incentive and monetary compensation [19]. As a result, diary studies are extremely difficult to implement, and only a handful of diary data sets has been available in the past decade [16-18]. Unlike other diary data sets, the data set we used for analysis integrated the aforementioned standardized QoL measurements in multiple waves, which appropriately facilitated our research goal.

Figure 1. Schematic plot of the data collection procedure. WHOQOL: World Health Organization Quality of Life survey.

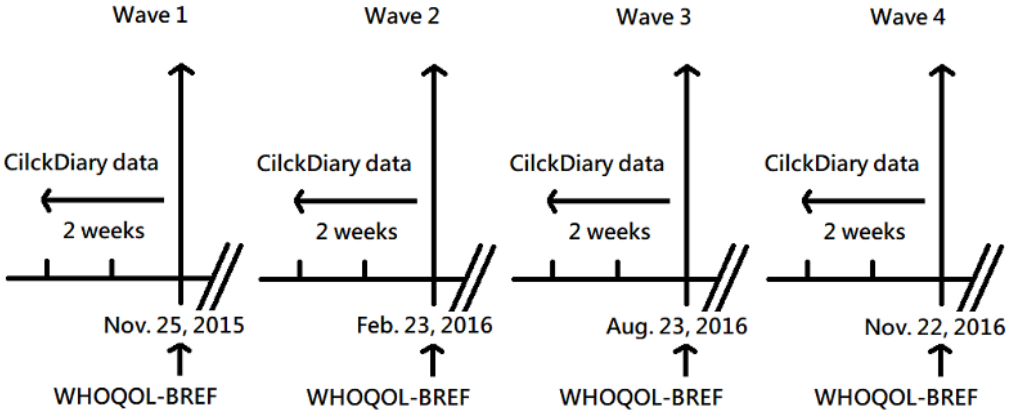


Table 1. Summary statistics of the ego-level attributes (n=149).

Variable	Results
Male gender, n (%)	39 (26.2)
Male gender, SD	44
Age (years), mean (SD)	35.9 (12.6)
Education[a], mean (SD)	3.99 (0.5)
Nonneurotic personality[b], mean (SD)	1.85 (0.71)
Exercise (hours), mean (SD)	0.27 (0.31)
Sleep well, n (%)	61 (40.9)
Sleep well, SD	49.2
BMI 21-24 kg/m^2, n (%)	52 (34.9)
BMI 21-24 kg/m^2, SD	47.7
Network size (log scale), mean (SD)	4.26 (0.71)

[a]Education was coded in 6 levels as follows: 0 = no education, 1 = elementary school, 2 = middle school, 3 = high school, 4 = college/university, and 5 = graduate school.

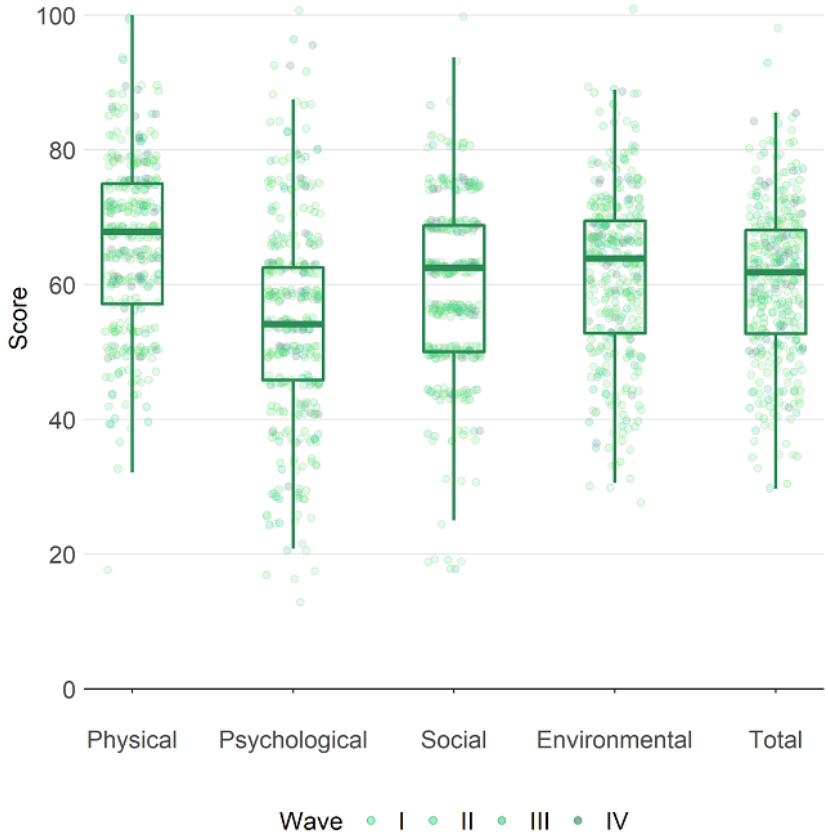
[b]Nonneurotic personality was measured with a scale ranging from 0 (disagree strongly) to 3 (agree strongly).

Response Variables

The WHOQOL-100 survey was developed through a collaboration of 15 sites around the world using a common protocol with different languages [24]. Each of the 100 items in this original scale contains responses on a 5-point Likert interval. To simplify the instrument, the WHOQOL-BREF [22] was developed. The WHOQOL-BREF consists of only 26 crosscultural items spanning 4 dimensions of QoL: physical, psychological, social, and environmental. With excellent reliability, this 26-item scale measures people's QoL across different cultural and societal settings [25,26]. In addition to the 26 items, 2 extra items are added to represent unique cultural attributes relevant to the QoL of Taiwanese people. The first concerns personal respectfulness, which belongs to the social dimension. The second focuses on the availability of food, part of the environmental dimension. From the 4-wave WHOQOL-BREF survey, we calculated scores for the physical, psychological, social, and environmental dimensions as follows. First, we divided the 28 (26 plus 2) items in WHOQOL-BREF into the 4 dimensions according to the WHOQOL-BREF guidelines [22]. To make the 4 scores comparable, we further rescaled the scores into a range between 0 and 100. We also calculated a summary score by summing the rescaled 4-dimensional scores. This summary score was further rescaled into a range between 0 and 100 for measuring a participant's overall QoL [22].

Figure 2 shows a boxplot of the 4 waves of the QoL scores for physical, psychological, social, environmental, and overall dimensions of the participants. The means (SDs) of the 5 QoL scores were 66.0 (12.4), 54.7 (15.5), 59.1 (12.7), 62.1 (12.4), and 60.5 (11.4), respectively.

Figure 2. Boxplot of quality of life scores, as calculated using the guideline described in the World Health Organization Quality of Life (WHOQOL) Taiwan Version [22]: physical, psychological, social, environmental, and corresponding total scores.



Contact Data

The ClickDiary platform is a simplified version of a contact diary that aims to reach more volunteer participants by reducing the burden of diary keeping while retaining the core elements of contact diaries. It allows participants to report information about daily contacts via a website with a responsive design adjusted for different electronic devices. Unlike conventional means of network data collection, ClickDiary contributes to the methodology with 2 distinct features. First, the platform is cost-effective since it pays exclusive attention to personal network information and does not require knowledge of the whole network [27]. Second, the diary design helps generate network data on a daily basis. The contact networks based on such diary data follow a longitudinal format that can yield more continuous information. This differs from both the network data collected via one-shot surveys and the unstructured contact records collected from social media.

By tracing all kinds of contact between ego and alters, ClickDiary probes the type, time, location, and duration of each contact. It also asks the participant to report the consequences of the contact, such as the extent to which each specific contact benefits ego and leaves ego in a good mood after the contact. Besides personal information about ego, ClickDiary also requires the participant to provide detailed information about each alter's demographic and socioeconomic background. Each participant further judges the strength of all ego-alter ties and estimates the strength of all alter-alter ties within each egocentric network. This further enables us to investigate the impact of the social network structure on a particular contact outcome. With a large number of alters, reporting this kind of information may take a long time and can be laboriously intensive, and some commitment is required to accomplish the job. To enhance this commitment, we designed a monetary incentive to encourage participants to report responsibly and correctly. The research team checked the data quality each week and set up rules to exclude unreliable data and participants.

Data collected via ClickDiary are organized in a hierarchical structure that involves both space-time and interpersonal dimensions. This hierarchical structure contains 3 levels: (1) ego level, at which the participant (ego) is identified; (2) alter level, at which an alter of the participant (ego) is identified; (3) contact level, at which a daily contact between the participant and an alter is identified. In our study, we explored the hierarchical structure to calculate the attributes of each ego and the attributes of each egocentric network. In other words, we used the data collected at the ego level to calculate personal attributes for the participant, and the data collected at the alter and contact levels were used to calculate the attributes of the social network surrounding the participant. In addition, we summarized each network attribute in a propensity score ranging between 0 and 1, which in turn served as a covariate (independent variable) in the subsequent regression analysis on how QoL is linked with social networks. Because we asked the participants about their QoL conditions during the previous 2

weeks, we calculated the propensity scores using only data collected 2 weeks before a wave of the WHOQOL-BREF survey was conducted. In particular, the propensity score of a social network attribute was calculated by aggregating the frequency of the contacts that matched this attribute during the previous 2 weeks.

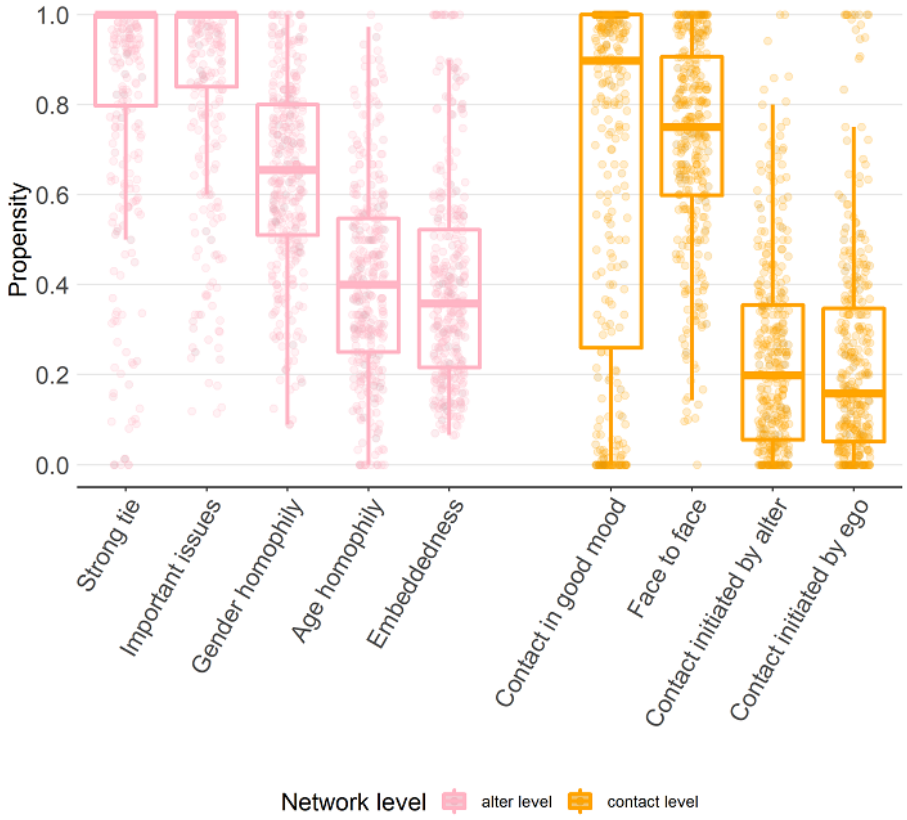
Independent Variables

The independent variables in our analysis cover ego's personal attributes and the attributes of the social network surrounding each ego. We utilized the following 8 personal attributes for each ego: gender, age, education, nonneurotic personality, daily exercise time, sleep quality, BMI, and personal network size (Table 1).

We used 5 social network attributes at the alter level for each ego. All 5 network attributes were summarized as propensity scores that were calculated using diary data collected during the 2 weeks before we queried respondents using the WHOQOL-BREF scale. The 5 attributes were the proportion of alters who were strongly tied with ego (strong tie); proportion of alters with whom ego could discuss important issues (important issue); proportion of alters of the same gender as ego (gender homophily); proportion of alters in the same age cohort as ego (age homophily); and averaged embeddedness score of the alters (embeddedness, or the extent to which alters knew one another within each egocentric network).

We further reconstructed 4 social network attributes for each ego from data at the contact level: proportion of the contacts in which ego felt that the alters were in a good mood, proportion of contacts that were conducted face-to-face, proportion of contacts that were initiated by the alters, and proportion of contacts that were initiated by ego. Figure 3 presents a boxplot of the 5 alter-level attributes and 4 contact-level attributes. Because these attributes are defined in terms of proportion, their values are between 0 and 1.

Figure 3. Boxplot of the alter-level attributes (pink) and contact-level attributes (orange).



Statistical Analysis

We investigated how QoL is linked to social networks by conducting statistical analysis using the 4-dimension scores and the total score of QoL as the response variables and the social network attributes as the independent variables. Here, we adopted a regression approach to analyzing the data. The regression approach is an effective way to investigate the relationship between a response variable and a set of independent variables. As we are interested in the relationship between QoL and the social networks, a regression model can help us to clarify what aspects of the social networks are associated with QoL. It can also control the variables that may exhibit confounding effects on the response. However, to conduct a proper regression analysis, we needed to pay attention to possible correlations among the independent variables. Figure 4 shows a plot of correlations of the independent variables: 8 ego-level attributes, 5 alter-level attributes, and 4 contact-level attributes. Because the correlations among these attributes were not high, these attributes were appropriate as independent

variables in the regression analysis. The weak correlations among the ego-level, alter-level, and contact-level attributes also allowed us to use a linear regression model to examine the relationship between the response variable (QoL measure) and independent variables (the attributes) since we did not need to consider collinearity among the independent variables. Collinearity among independent variables will destabilize computation of the Gram matrix of independent variables. It will increase the variance of the estimated regression coefficients, further distorting interpretation of the results provided by the regression analysis. On the other hand, because the response variable was measured repeatedly, the measured scores of the consecutive waves tended to correlate with each other. Figure 5 shows correlation plots of the 4-dimension scores and the total QoL score in the 4-wave survey (labeled I, II, III, and IV). These plots indicate that the scores of consecutive waves were highly correlated. These high correlations may have had a serious impact on regression model estimation, particularly on the variance of the estimated regression coefficients. To address this problem, we adopted the generalized estimating equation (GEE) method to estimate the regression models [28]. The GEE method takes the correlation structure of the response into account when estimating the coefficients of the regression models.

Figure 4. Correlation plot of the ego-level, alter-level, and contact-level attributes.

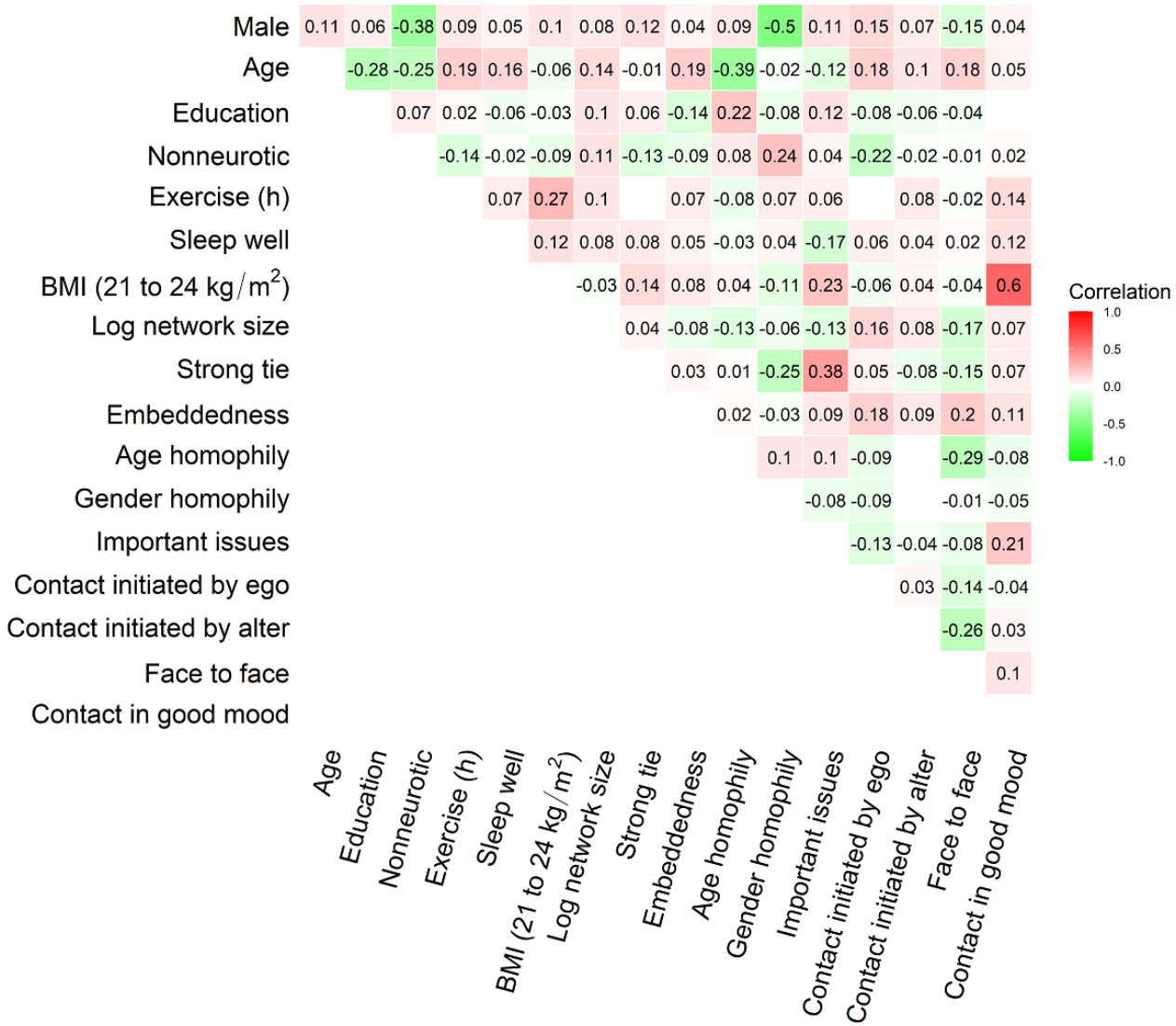
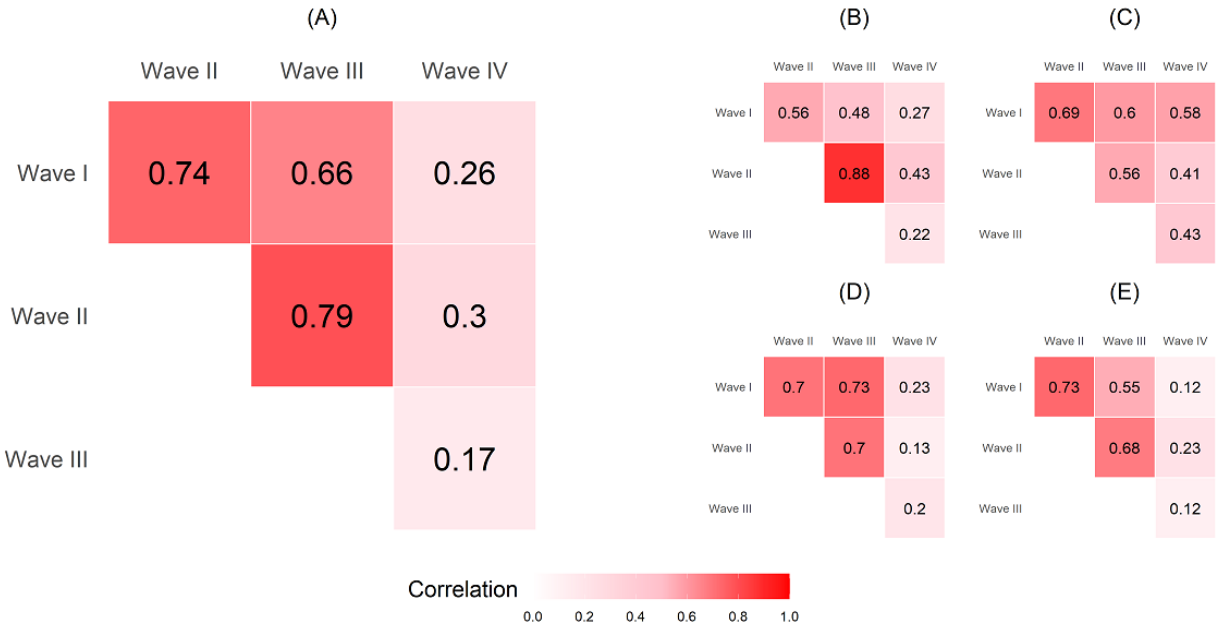


Figure 5. Correlation plot of the responses in the 4-wave surveys: (A) total score, (B) physical score, (C) psychological score, (D) social score, (E) environmental score.



Results

We estimated the regression models separately for each of the 4 dimensions of QoL scores and the total score. Figure 6 shows the estimation results for the model with the total score as the response variable. The results revealed that a participant's self-reported total score of QoL was higher when the daily contacts consisted of a larger proportion of alters who were strongly tied (estimated regression coefficient 7.15, 95% CI 1.88 to 12.42), who could discuss important issues with the participant (coefficient 7.52, 95% CI 0.30 to 14.75), or who were in a good mood (coefficient 5.02, 95% CI 1.45 to 8.58) as well as contact via face-to-face meetings (coefficient 5.57, 95% CI 0.46 to 10.67). In addition, total QoL score was higher if contact was initiated by alters (coefficient 3.99, 95% CI 7.42 to 0.57). On the other hand, we also found that the size of the social network surrounding a participant did not have an impact on the participant's total score (coefficient –0.241, 95% CI –2.50 to 2.02). Although contacting a high proportion of alters who are embedded in a participant's personal network may have a negative impact on the participant's total QoL score, such a structural factor was not statistically significant (coefficient –6.91, 95% CI –14.02 to 0.20).

Figure 7 shows that a participant's physical QoL score was higher when more contact was initiated by alters (coefficient 7.88, 95% CI 2.75 to 13.01). Figure 8 further indicates that a participant's psychological QoL score was positively associated with the proportion of contact that was face-to-face (coefficient 6.88, 95% CI 0.40 to 13.36) and contact during which the alter was in a good mood (coefficient 8.59, 95% CI 3.71 to 8.59). Psychological QoL also seemed to be better with a higher proportion of contact in which the alter was strongly tied to the participant (coefficient 8.86, 95% CI 1.33 to 16.39). Figure 9 shows that a participant's social QoL score was positively associated with the proportion of contact in which the alter was in a good mood (coefficient 6.57, 95% CI 2.68 to 10.46), the alter could discuss important issues (coefficient 9.02, 95% CI 0.98 to 17.05), and the alter was strongly tied with the participant (coefficient 9.17; 95% CI 2.56 to 15.78). Figure 10 suggests that a participant's environmental score was positively associated with the proportion of alters who could discuss important issues (coefficient 10.01, 95% CI 2.23 to 17.79) and the proportion of alters who were strongly tied with the participant (coefficient 8.02, 95% CI 1.60 to 14.44) but negatively associated with the embeddedness score (coefficient –13.58, 95% CI –21.66 to –5.50).

Figure 6. Estimation results from the regression model with the total score as the response (dependent variable). The estimated regression coefficients corresponding to the contact-level attributes, alter-level attributes, and ego-level attributes are shown in orange, pink, and light blue, respectively, and the bars represent the 95% CI for the estimated regression coefficient.

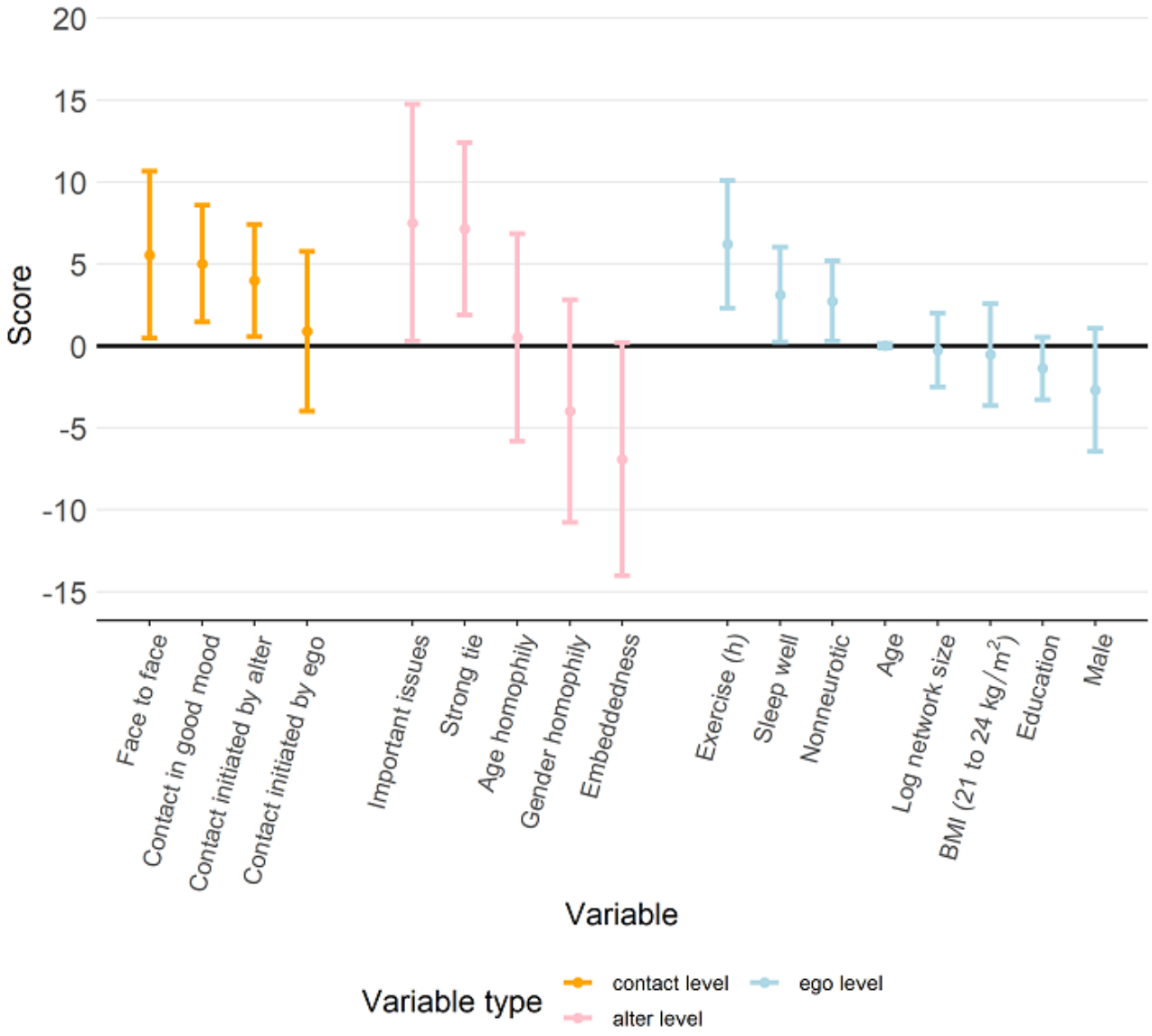


Figure 7. Estimation results from the regression model with the physical score as the response (dependent variable). The estimated regression coefficients corresponding to the contact-level attributes, alter-level attributes, and ego-level attributes are shown in orange, pink, and light blue, respectively, and the bars represent the 95% CI for the estimated regression coefficient.

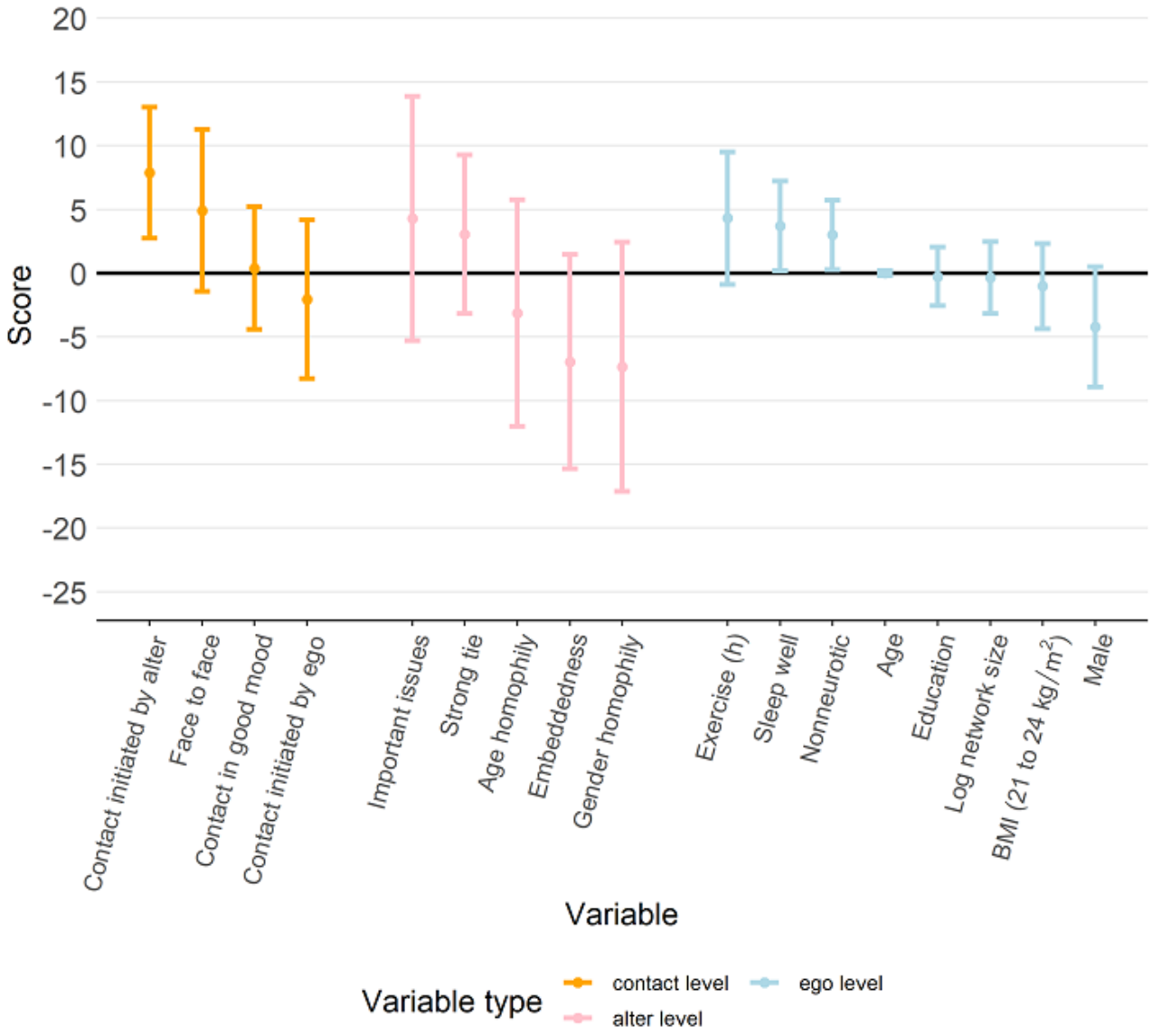


Figure 8. Estimation results from the regression model with the psychological score as the response (dependent variable). The estimated regression coefficients corresponding to the contact-level attributes, alter-level attributes, and ego-level attributes are shown in orange, pink, and light blue, respectively, and the bars represent the 95% CI for the estimated regression coefficient.

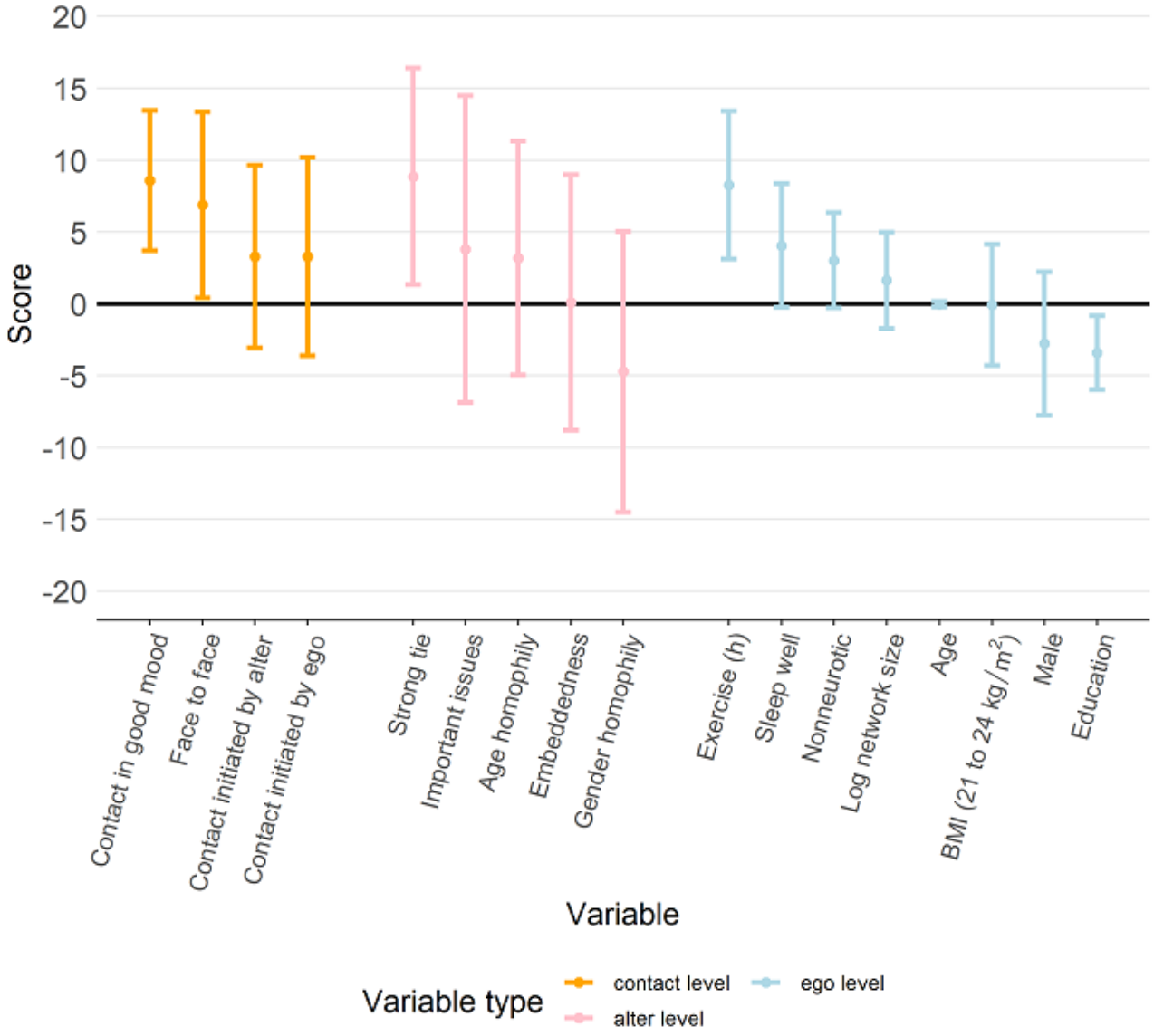


Figure 9. Estimation results from the regression model with the social score as the response (dependent variable). The estimated regression coefficients corresponding to the contact-level attributes, alter-level attributes, and ego-level attributes are shown in orange, pink, and light blue, respectively, and the bars represent the 95% CI for the estimated regression coefficient.

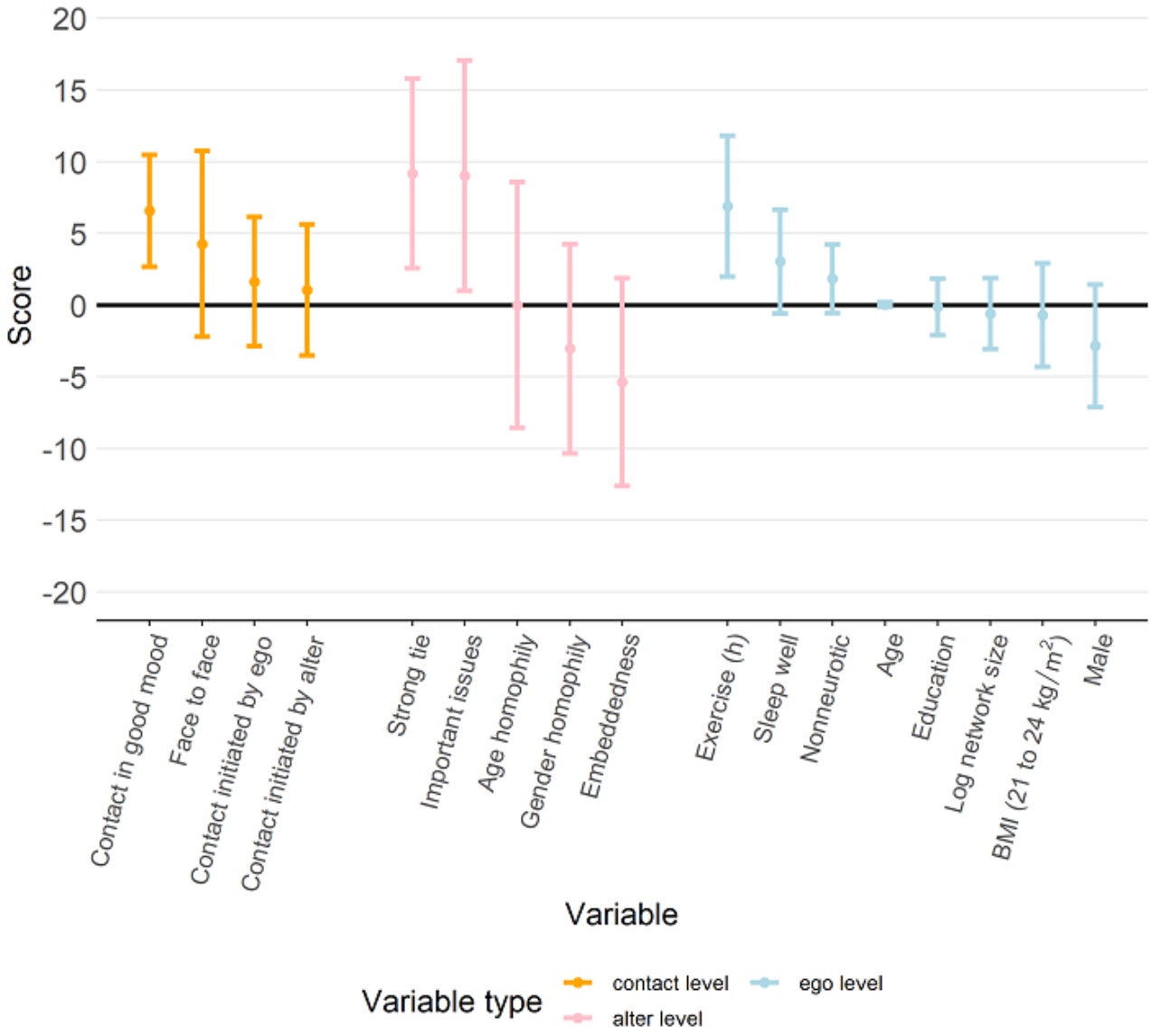
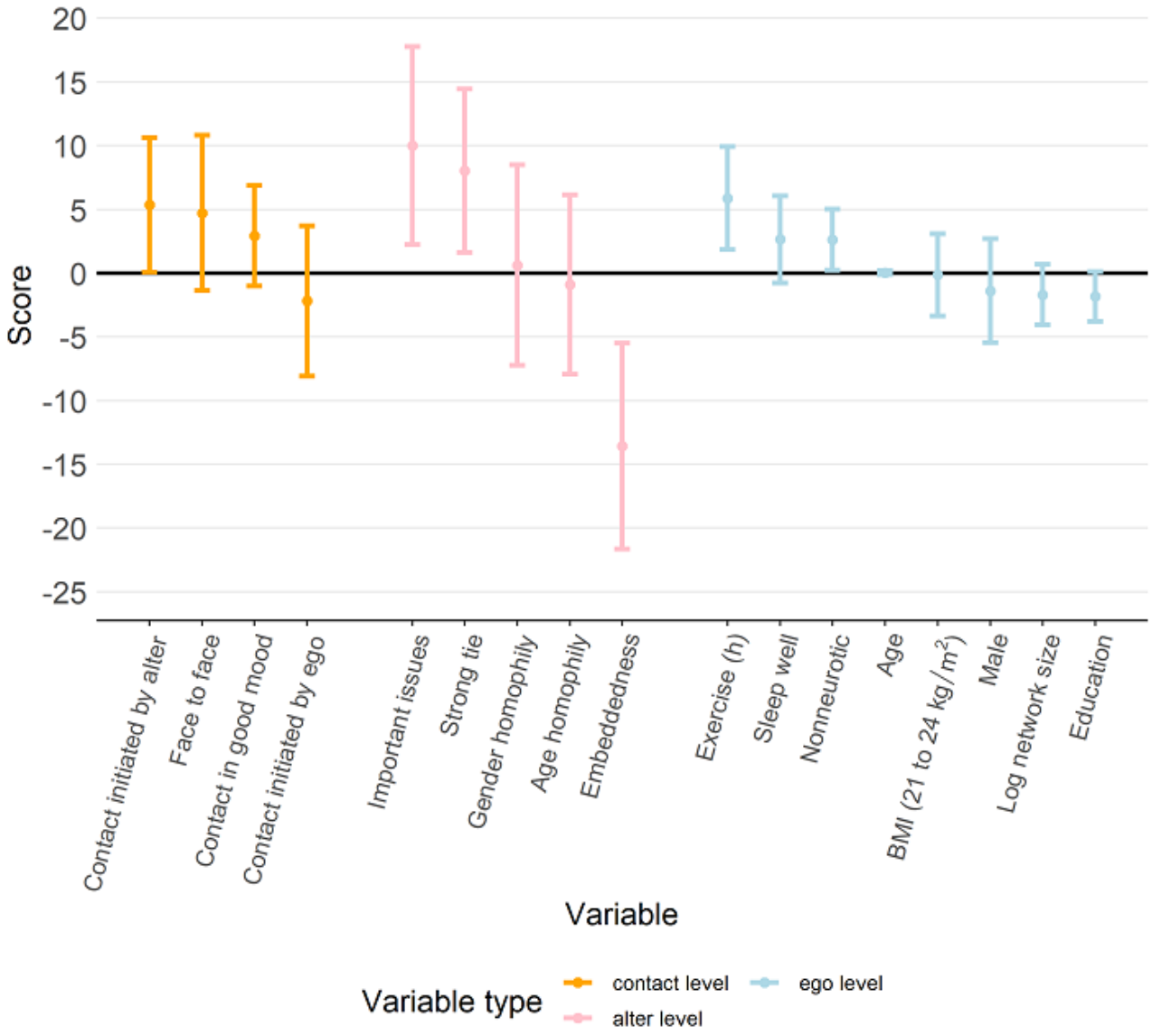


Figure 10. Estimation results from the regression model with the environmental score as the response (dependent variable). The estimated regression coefficients corresponding to the contact-level attributes, alter-level attributes, and ego-level attributes are shown in orange, pink, and light blue, respectively, and the bars represent the 95% CI for the estimated regression coefficient.



Discussion

In this paper, we investigated the relationships between QoL and social networks by jointly analyzing 2 sets of online platform data collected via longitudinal surveys and contact diaries. According to the regression analysis, the overall QoL among these healthy young adults tended to be higher when they contacted those who were connected well with them, who could discuss important issues with them, and who made them feel better during an interaction. These findings imply that both functional and structural aspects of social networks play important roles in shaping one's QoL. The participants tended to benefit by interacting with people they knew well, who were trustworthy, and who were in a positive mood. Overall QoL was also more likely to increase if more of their daily contacts were initiated by the other party. This interesting finding may be surprising because earlier diary studies indicated that a person tends to benefit from a specific social exchange while initiating a contact [29]. Overall, however, constantly reaching out to others may not bring one the best QoL in the long run. Being approached or "invited" by others, after all, may signal a form of "respectfulness" embedded in social relationships that helps boost one's perception of QoL.

In contrast, network size alone was not significantly linked to overall QoL, after taking both the functional and structural aspects of social networks into account. This finding is not surprising because network size was used as a proxy measure for functional and structural aspects of social networks. Once the functional and structural aspects of the social network are

both incorporated into the model, the effect of network size disappears.

More specifically, the functional and structural aspects of social networks may play various roles in shaping different dimensions of QoL. For example, the participants' physical QoL increased when their daily contacts were frequently initiated by other people, even though this factor did not help other dimensions of QoL. In addition to the possible "respectfulness" suggested in the previous paragraph, this finding also may have to do with one's physical condition: Better physical health tends to facilitate or attract more invitations, or initiations, from social contacts. Moreover, the psychological and social dimensions of people's QoL may get better if they frequently interact with those who have positive moods. However, this "positive energy" factor does not have a significant impact on helping the physical and environmental dimensions of QoL.

There are noted limitations to this study. Although our data follow a longitudinal format, the results cannot be used to verify the causal relationships between social networks and QoL. In addition, our results are based on information about the social network surrounding the participant, which has been constructed through contact diaries repeatedly recorded by the participant. Although this diary approach tends to yield active and comprehensive archives of personal networks and is thus more stable and reliable than one-shot survey data, these self-reported contact records may not be free from recall biases. It is thus difficult to rule out potential measurement errors, even though we tried to verify any doubtful information by calling participants to double check and correct the data.

Our data have a repeated but "imbalanced" measurement format in which participants did not have equal observations, which may be another concern for statistical analysis. Because the social network effects we were interested in were neither individual-specific nor time-specific, we could still estimate our regression models by treating the data as repeated measurement data without a strong assumption on the missing mechanism of the unobserved data points. What we needed to pay attention to was the within-subject correlations between each measurement, which were high in our case, as shown in Figure 6. We tackled this issue by estimating our regression model with the GEE method, a method developed for dealing with data in which observations are dependent.

In our regression analysis, we treated only social network attributes as the main factors. Such network effects may interact with demographic and socioeconomic factors; recent studies have found that social networks play an important role in mediating the impacts of gender [30,31], age [32], and ethnicity [33] on health-related QoL. Not only does the social network function as a mediator but it may also reciprocally influence other factors in their impacts on QoL [34]. To take these possible multiple roles into account, future studies may want to examine the circumstances under which the social network attributes mediate the sociodemographic covariates that control the paths between other factors and QoL.

Conflicts of Interest

None declared.

References

1. Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. Am J Epidemiol 1979 Mar;109(2):186-204. [doi: 10.1093/oxfordjournals.aje.a112674] [Medline: 425958]
2. House JS, Landis KR, Umberson D. Social relationships and health. Science 1988 Jul 29;241(4865):540-545. [doi: 10.1126/science.3399889] [Medline: 3399889]
3. Cohen S, Doyle WJ, Skoner DP, Rabin BS, Gwaltney JM. Social ties and susceptibility to the common cold. JAMA 1997 Jun 25;277(24):1940-1944. [Medline: 9200634]
4. Michael YL, Berkman LF, Colditz GA, Holmes MD, Kawachi I. Social networks and health-related quality of life in breast cancer survivors: a prospective study. J Psychosom Res 2002 May;52(5):285-293. [doi: 10.1016/s0022-3999(01)00270-7] [Medline: 12023125]
5. Helgeson VS. Social support and quality of life. Qual Life Res 2003;12 Suppl 1:25-31. [doi: 10.1023/a:1023509117524] [Medline: 12803308]
6. Rüesch P, Graf J, Meyer PC, Rössler W, Hell D. Occupation, social support and quality of life in persons with schizophrenic or affective disorders. Soc Psychiatry Psychiatr Epidemiol 2004 Sep;39(9):686-694. [doi: 10.1007/s00127-004-0812-y] [Medline: 15672288]
7. Christakis NA, Fowler JH. The spread of obesity in a large social network over 32 years. N Engl J Med 2007 Jul 26;357(4):370-379. [doi: 10.1056/NEJMsa066082] [Medline: 17652652]
8. Fowler JH, Christakis NA. Dynamic spread of happiness in a large social network: longitudinal analysis over 20 years in the Framingham Heart Study. BMJ 2008 Dec 04;337:a2338 [FREE Full text] [doi: 10.1136/bmj.a2338] [Medline: 19056788]
9. Christakis NA, Fowler JH. The collective dynamics of smoking in a large social network. N Engl J Med 2008 May 22;358(21):2249-2258 [FREE Full text] [doi: 10.1056/NEJMsa0706154] [Medline: 18499567]
10. Cornwell B, Laumann EO. Network position and sexual dysfunction: implications of partner betweenness for men. AJS 2011 Jul;117(1):172-208 [FREE Full text] [doi: 10.1086/661079] [Medline: 22003520]

11. Masedu F, Mazza M, Di Giovanni C, Calvarese A, Tiberti S, Sconci V, et al. Facebook, quality of life, and mental health outcomes in post-disaster urban environments: the l'aquila earthquake experience. Front Public Health 2014;2:286 [FREE Full text] [doi: 10.3389/fpubh.2014.00286] [Medline: 25566527]
12. Achat H, Kawachi I, Levine S, Berkey C, Coakley E, Colditz G. Social networks, stress and health-related quality of life. Qual Life Res 1998 Dec;7(8):735-750. [doi: 10.1023/a:1008837002431] [Medline: 10097622]
13. Hemmati A, Chung KSK. Social networks and quality of life: The national health interview survey. 2014 Presented at: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014); August 17-20, 2014; Beijing, China. [doi: 10.1109/asonam.2014.6921644]
14. Lim JT, Park J, Lee J, Oh J, Kim Y. The relationship between the social network of community-living elders and their health-related quality of life in Korean province. J Prev Med Public Health 2013 Jan;46(1):28-38 [FREE Full text] [doi: 10.3961/jpmph.2013.46.1.28] [Medline: 23407568]
15. Perry BL, Pescosolido BA. Social network activation: the role of health discussion partners in recovery from mental illness. Soc Sci Med 2015 Jan;125:116-128 [FREE Full text] [doi: 10.1016/j.socscimed.2013.12.033] [Medline: 24525260]
16. Chan T, Yen T, Fu Y, Hwang J. ClickDiary: Online Tracking of Health Behaviors and Mood. J Med Internet Res 2015 Jun 15;17(6):e147 [FREE Full text] [doi: 10.2196/jmir.4315] [Medline: 26076583]
17. Chan T, Yen T, Hu T, Fu Y, Hwang J. Detecting concurrent mood in daily contact networks: an online participatory cohort study with a diary approach. BMJ Open 2018 Jul 10;8(7):e020600 [FREE Full text] [doi: 10.1136/bmjopen-2017-020600] [Medline: 29991627]
18. Fu Y, Velema T, Hwang J. Upward contacts in everyday life: Benefits of reaching hierarchical relations in ego-centered networks. Social Networks 2018 Jul;54:266-278 [FREE Full text] [doi: 10.1016/j.socnet.2018.03.002]
19. Fu Y. Contact diaries. Field Methods 2016 Jul 21;19(2):194-217. [doi: 10.1177/1525822x06298590]
20. Kahneman D, Krueger AB, Schkade DA, Schwarz N, Stone AA. A survey method for characterizing daily life experience: the day reconstruction method. Science 2004 Dec 03;306(5702):1776-1780. [doi: 10.1126/science.1103572] [Medline: 15576620]
21. The WHOQOL Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med 1998 May;28(3):551-558. [doi: 10.1017/s0033291798006667] [Medline: 9626712]
22. WHOQOL- BVG. The guideline of the WHOQOL-BREF Taiwan Version. The guideline of the WHOQOL-BREF Taiwan Version 2000 Jan 01:2000. [doi: 10.1007/978-94-007-0753-5_104548]
23. Yao K, Chung CW, Yu CF, Wang JD. Development and applications of the WHOQOL-Taiwan version. Formosan Journal of Medicine 2002;101:342-351 [FREE Full text]
24. The WHOQOL Group. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. Soc Sci Med 1998 Jun;46(12):1569-1585. [doi: 10.1016/s0277-9536(98)00009-4] [Medline: 9672396]
25. Skevington SM, Lotfy M, O'Connell KA, WHOQOL Group. The World Health Organization's WHOQOL-BREF quality of life assessment: psychometric properties and results of the international field trial. A report from the WHOQOL group. Qual Life Res 2004 Mar;13(2):299-310. [doi: 10.1023/B:QURE.0000018486.91360.00] [Medline: 15085902]
26. Berlim MT, Pavanello DP, Caldieraro MAK, Fleck MPA. Reliability and validity of the WHOQOL BREF in a sample of Brazilian outpatients with major depression. Qual Life Res 2005 Mar;14(2):561-564. [doi: 10.1007/s11136-004-4694-y] [Medline: 15892446]
27. Cohen R, Havlin S, Ben-Avraham D. Efficient immunization strategies for computer networks and populations. Phys Rev Lett 2003 Dec 12;91(24):247901. [doi: 10.1103/PhysRevLett.91.247901] [Medline: 14683159]
28. Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. Biometrika 1986;73(1):13-22. [doi: 10.1093/biomet/73.1.13]
29. Fu Y, Ho H, Chen H. Weak ties and contact initiation in everyday life: Exploring contextual variations from contact diaries. Social Networks 2013 Jul;35(3):279-287 [FREE Full text] [doi: 10.1016/j.socnet.2013.02.004]
30. García EL, Banegas JR, Pérez-Regadera AG, Cabrera RH, Rodríguez-Artalejo F. Social network and health-related quality of life in older adults: a population-based study in Spain. Qual Life Res 2005 Mar;14(2):511-520. [doi: 10.1007/s11136-004-5329-z] [Medline: 15892440]
31. Tobiasz-Adamczyk B, Galas A, Zawisza K, Chatterji S, Haro JM, Ayuso-Mateos JL, et al. Gender-related differences in the multi-pathway effect of social determinants on quality of life in older age-the COURAGE in Europe project. Qual Life Res 2017 Jul;26(7):1865-1878 [FREE Full text] [doi: 10.1007/s11136-017-1530-8] [Medline: 28258420]
32. Levula A, Wilson A, Harré M. The association between social network factors and mental health at different life stages. Qual Life Res 2016 Jul 15;25(7):1725-1733. [doi: 10.1007/s11136-015-1200-7] [Medline: 26669317]
33. Bergeron G, Lundy De La Cruz N, Gould LH, Liu SY, Levanon Seligson A. Association between racial discrimination and health-related quality of life and the impact of social relationships. Qual Life Res 2020 Oct;29(10):2793-2805 [FREE Full text] [doi: 10.1007/s11136-020-02525-2] [Medline: 32444931]
34. Cheng GH, Malhotra R, Chan A, Østbye T, Lo JC. Weak social networks and restless sleep interrelate through depressed mood among elderly. Qual Life Res 2018 Oct;27(10):2517-2524. [doi: 10.1007/s11136-018-1895-3] [Medline: 29869296]

Abbreviations

GEE: generalized estimating equation
QoL: quality of life
WHOQOL: World Health Organization Quality of Life

Edited by G Eysenbach; submitted 22.08.20; peer-reviewed by R Poluru, M Reblin; comments to author 29.09.20; revised version received 24.11.20; accepted 10.12.21; published 28.01.22.

Please cite as:
Yen TJ, Chan TC, Fu YC, Hwang JS
Quality of Life and Multilevel Contact Network Structures Among Healthy Adults in Taiwan: Online Participatory Cohort Study
J Med Internet Res 2022;24(1):e23762
URL: https://www.jmir.org/2022/1/e23762
doi:10.2196/23762
PMID:35089142

Original Paper

Describing Transitions in Adherence to Physical Activity Self-monitoring and Goal Attainment in an Online Behavioral Weight Loss Program: Secondary Analysis of a Randomized Controlled Trial

Melissa L Stansbury[1], PhD; Jean Harvey[2], PhD; Rebecca A Krukowski[3], PhD; Christine A Pellegrini[1], PhD; Xuewen Wang[1], PhD; Delia Smith West[1], PhD

[1]Department of Exercise Science, Arnold School of Public Health, University of South Carolina, Columbia, SC, United States

[2]Department of Nutrition and Food Sciences, College of Agriculture and Life Sciences, University of Vermont, Burlington, VT, United States

[3]Department of Public Health Sciences, School of Medicine, University of Virginia, Charlottesville, VA, United States

Corresponding Author:
Melissa L Stansbury, PhD
Department of Exercise Science
Arnold School of Public Health
University of South Carolina
Discovery 1 Bldg, Suite 403
915 Greene Street
Columbia, SC, 29208
United States
Phone: 1 803 777 0510
Fax: 1 803 777 5457
Email: stansbm@email.sc.edu

Abstract

Background: Standard behavioral weight loss interventions often set uniform physical activity (PA) goals and promote PA self-monitoring; however, adherence remains a challenge, and recommendations may not accommodate all individuals. Identifying patterns of PA goal attainment and self-monitoring behavior will offer a deeper understanding of how individuals adhere to different types of commonly prescribed PA recommendations (ie, minutes of moderate-to-vigorous physical activity [MVPA] and daily steps) and guide future recommendations for improved intervention effectiveness.

Objective: This study examined weekly patterns of adherence to step-based and minute-based PA goals and self-monitoring behavior during a 6-month online behavioral weight loss intervention.

Methods: Participants were prescribed weekly PA goals for steps (7000-10,000 steps/day) and minutes of MVPA (50-200 minutes/week) as part of a lifestyle program. Goals gradually increased during the initial 2 months, followed by 4 months of fixed goals. PA was self-reported daily on the study website. For each week, participants were categorized as adherent if they self-monitored their PA and met the program PA goal, suboptimally adherent if they self-monitored but did not meet the program goal, or nonadherent if they did not self-monitor. The probability of transitioning into a less adherent status was examined using multinomial logistic regression.

Results: Participants (N=212) were predominantly middle-aged females with obesity, and 67 (31.6%) self-identified as a racial/ethnic minority. Initially, 73 (34.4%) participants were categorized as adherent to step-based goals, with 110 [51.9%] suboptimally adherent and 29 [13.7%] nonadherent, and there was a high probability of either remaining suboptimally adherent from week to week or transitioning to a nonadherent status. However, 149 (70.3%) participants started out adherent to minute-based goals (34 [16%] suboptimally adherent and 29 [13.7%] nonadherent), with suboptimally adherent seen as the most variable status. During the graded goal phase, participants were more likely to transition to a less adherent status for minute-based goals (odds ratio [OR] 1.39, 95% CI 1.31-1.48) compared to step-based goals (OR 1.24, 95% CI 1.17-1.30); however, no differences were seen during the fixed goal phase (minute-based goals: OR 1.06, 95% CI 1.05-1.08; step-based goals: OR 1.07, 95% CI 1.05-1.08).

Conclusions: States of vulnerability to poor PA adherence can emerge rapidly and early in obesity treatment. There is a window of opportunity within the initial 2 months to bring more people toward adherent behavior, especially those who fail to meet the

prescribed goals but engage in self-monitoring. Although this study describes the probability of adhering to step- and minute-based targets, it will be prudent to determine how individual characteristics and contextual states relate to these behavioral patterns, which can inform how best to adapt interventions.

Trial Registration: ClinicalTrials.gov NCT02688621; https://clinicaltrials.gov/ct2/show/NCT02688621

(J Med Internet Res 2022;24(1):e30673) doi:10.2196/30673

KEYWORDS

physical activity; adherence; self-monitoring; goal attainment; lifestyle intervention

Introduction

Adherence to physical activity (PA) is a pervasive challenge in behavioral weight control programs. Many individuals fail to achieve program recommendations and adopt a physically active lifestyle [1,2]. As a result, these individuals do not obtain the multitude of immediate and long-term health benefits associated with engaging in PA, including chronic disease prevention and weight management [3,4]. A deeper understanding of the underlying pattern of an individual's adherence to program recommendations may reveal information about what contributes to deterioration in PA adherence and indicate important targets for intervention.

Ordinarily, all individuals receive the same programmatic goals, which gradually progress to targeted PA levels [5-7]. Recommendations are most often provided for moderate-to-vigorous intensity physical activity (MVPA) in terms of minute-based goals, such as 150 minutes/week of MVPA [8]. More recently, step-based goals (eg, 10,000 steps/day) have also been incorporated as a means to bolster the accumulation of activity throughout the day. Individuals who achieve program MVPA goals during treatment tend to lose more weight and are more likely to maintain high levels of activity in the future [1,9-12]. However, the stability in adherence to step- and minute-based goals over time, as well as differences in adherence between these 2 types of goal prescriptions, have yet to be identified. Responsiveness to program recommendations, along with a more precise description of how the type and timing of PA goals impact adherence, can uncover key areas for consideration when prescribing PA for behavioral weight control and guide intervention tailoring for those at risk of poor adherence.

Self-monitoring, a key element of lifestyle programs intended to promote behavioral awareness and fulfillment of recommendations [4,13], serves as a reliable indicator of weight loss and PA engagement [10,14]. Failure to initiate self-monitoring of MVPA and inconsistent recording of MVPA are associated with weight gain and lower activity levels during a behavioral weight loss program [15]. Whether self-monitoring of daily steps follows a similar pattern to MVPA remains to be determined. Describing how individuals adhere to self-monitoring of different types of program PA goals from one week to the next may point toward which type of metric (steps or minutes of MVPA) individuals are more likely to engage in during a behavioral weight control program.

Others have examined transitions in levels of adherence to dietary self-monitoring and found that individuals who partially self-monitor in 1 week have a greater chance of improving to full self-monitoring during the first 2 months of treatment, whereas after the first 2 months, partial self-monitoring in 1 week foreshadows discontinuation of self-monitoring [16]. Using a similar approach to examining PA will indicate whether there are meaningful patterns of PA self-monitoring and engagement, as well as when these patterns emerge. In turn, potential targets for adaptive interventions may be identified, which could ultimately improve adherence in behavioral weight control programs and enhance outcomes for both PA promotion and weight management.

Examination of weekly adherence to the different types of PA goals offers insight into goal operationalization, progression, and timing, since lifestyle programs tend to prescribe goals on a weekly basis. Therefore, this study aims to describe weekly patterns of adherence to (1) step-based goals and self-monitoring steps and (2) minute-based goals and self-monitoring minutes of MVPA, as well as examine whether weekly transitional patterns differ between step-based goals versus minute-based goals, across 6 months of an online, group-based behavioral weight control program.

Methods

Study Design

This study was a secondary analysis of a randomized controlled trial (RCT) that evaluated the impact of financial incentives on weight loss in a group-based behavioral weight control program delivered online [17]. Briefly, participants received either (1) the online program augmented with financial incentives for achieving behavioral and weight loss goals or (2) the online program alone (ie, control group). Eligible participants had a body mass index (BMI) of 25-50 kg/m^2; were at least 18 years of age with access to a smartphone, computer, and the internet; had no medical contraindications to moderate-intensity PA (eg, brisk walking); and demonstrated the capacity to monitor their PA using their smartphone or personal fitness tracker prior to randomization. The study was approved by the institutional review boards at both clinical sites, and written informed consent was obtained from all participants. The investigation included only the control group (N=212) to avoid the potential of financial incentives obscuring PA adherence patterns (see Multimedia Appendix 1).

Program Overview

All participants received a goal-oriented behavioral weight control program delivered online, which has been shown effective in previous studies [18,19]. Weekly online group

sessions occurred synchronously via text-based chat led by an experienced facilitator for 6 months. Session topics, such as goal setting, problem solving, action planning, and relapse prevention, were supplemented with online materials, activities, and other resources on the study website to reinforce key strategies for self-regulation. A program goal of 10% weight loss was recommended through calorie reduction and increased PA. Participants were provided a dietary goal of 1200-1800 calories/day (≤25% from fat) based on their initial body weight. The same step- and minute-based PA goals were prescribed concurrently to all participants from weeks 3 to 24 in 2 phases. The graded goal phase began at week 3, with targets for 7000 steps/day and 50 minutes/week of MVPA; both goals progressed incrementally, reaching 10,000 steps/day and 150 minutes/week of MVPA at week 8. The fixed goal phase consisted of weeks 9-24, when both program PA goals were held constant at 10,000 steps/day and 200 minutes/week of MVPA. The MVPA goal was chosen to be consistent with the Centers for Disease Control and Prevention and American College of Sports Medicine guidelines [8]. Although there is much debate about the number of recommended steps, the 10,000-step goal is commonly used [20].

Participants were instructed to use their personal fitness tracker or a smartphone app to track their total steps and minutes of planned exercise of at least a moderate-to-vigorous intensity and then record both their total steps and minutes of MVPA on the study website daily. In other words, participants were asked to report both metrics for each day of the study period, as opposed to reporting only 1 metric, even if they did not perform any PA on that day. Self-monitoring records were reviewed by the facilitator, who provided tailored feedback via email on a weekly basis with positive reinforcement of healthy lifestyle choices, constructive guidance promoting potential areas for change, and reminders of the program recommendations.

Measures

Physical Activity

The total number of steps and minutes of MVPA both were self-reported by participants daily on the study website. Weekly totals for steps and for minutes of MVPA were calculated by summing the daily reported values, with each week beginning on the day of the scheduled group session.

Adherence to PA Goals

Adherence to the program step goals (yes/no) and the MVPA minute goal (yes/no) was separately determined for each participant and for each week of the program based on the proportion of the program goal achieved. For example, a participant could be categorized as nonadherent to the step goal and adherent to the MVPA goal. For steps, the average daily step count for a given week was divided by that week's program step goal and multiplied by 100%. For minutes of MVPA, the total minutes reported in a given week were divided by the program MVPA goal and multiplied by 100%. Proportions of ≥100% indicated that the corresponding goal was met (yes), and proportions of <100% indicated that the goal was unmet (no). A value of 0 was assumed for missing and implausible entries, including daily step counts of <1000 or >30,000 [21,22] and >1080 MVPA minutes/day [23,24].

Adherence to PA Self-Monitoring

Adherence to self-monitoring of steps (yes/no) and minutes of MVPA (yes/no) was based on whether an entry was submitted and calculated separately for each type of PA goal (ie, a participant could be categorized as adherent to 1 of the goals and not adherent to the other). For steps, a participant was considered to have self-monitored (yes) in a given week if ≥1 record of any step count was submitted, including a value of 0. Likewise, a participant was considered to have self-monitored minutes of MVPA (yes) for a given week if ≥1 record of any number of minutes, including 0, was submitted. If no entry was submitted for the corresponding PA goal on any of the 7 days, the participant was considered nonadherent to self-monitoring (no) for that week. The approach described here is consistent with previous work, which categorized a participant as adherent if ≥1 record of any value of exercise minutes was submitted for a given week and nonadherent if no records were submitted [25].

Categorization of PA Adherence

The variables *adherence to PA goal* and *adherence to PA self-monitoring* were used to categorize participants into 3 mutually exclusive categories each week for step-based goals and separately for minute-based goals. Participants were categorized as (1) *adherent* if they met the goal and self-monitored, (2) *suboptimally adherent* if they did not meet the goal but did self-monitor, and (3) *nonadherent* if they did not self-monitor (Table 1). It was assumed that those who did not self-monitor also did not meet the PA goal for that week. To the best of our knowledge, thresholds for adherence to PA goals and self-monitoring during behavioral weight control have not been standardized. Therefore, the categories were selected to mirror the established PA classifications outlined in the national guidelines (ie, *active/highly active*, *insufficiently active*, and *inactive*) [26], which are related to the degree of health benefits obtained [8]. Categorization of PA adherence to step-based goals did not influence how a participant was categorized for minute-based goals, and vice versa.

Table 1. Categorization of PA[a] adherence based on weekly program PA goals and PA self-monitoring.

Week	Adherent	Suboptimally adherent	Nonadherent
Steps			
3	≥7000 steps (≥100%) + ≥1 day of self-monitoring	1-6999 steps (0%<x<100%) + ≥1 day of self-monitoring	No self-monitoring
4-5	≥8000 steps (≥100%) + ≥1 day of self-monitoring	1-7999 steps (0%<x<100%) + ≥1 day of self-monitoring	No self-monitoring
6-7	≥9000 steps (≥100%) + ≥1 day of self-monitoring	1-8999 steps (0%<x<100%) + ≥1 day of self-monitoring	No self-monitoring
8-24	≥10,000 steps (≥100%) + ≥1 day of self-monitoring	1-9999 steps (0%<x<100%) + ≥1 day of self-monitoring	No self-monitoring
Minutes of MVPA[b]			
3-4	≥50 minutes (≥100%) + ≥1 day of self-monitoring	1-49 minutes (0%<x<100%) + ≥1 day of self-monitoring	No self-monitoring
5-6	≥100 minutes (≥100%) + ≥1 day of self-monitoring	1-99 minutes (0%<x<100%) + ≥1 day of self-monitoring	No self-monitoring
7-8	≥150 minutes (≥100%) + ≥1 day of self-monitoring	1-149 minutes (0%<x<100%) + ≥1 day of self-monitoring	No self-monitoring
9-24	≥200 minutes (≥100%) + ≥1 day of self-monitoring	1-199 minutes (0%<x<100%) + ≥1 day of self-monitoring	No self-monitoring

[a]PA: physical activity.

[b]MVPA: moderate-to-vigorous physical activity.

Statistical Analysis

Descriptive analyses were conducted to characterize the sample and determine frequencies for the number of participants meeting PA goals and self-monitoring their PA each week. Marginal proportions of weekly PA adherence categories were separately tabulated for each type of PA recommendation (steps and minutes of MVPA) beginning at week 3 and continuing through the end of the 6-month period. The weekly transition probabilities of adherence status were then independently summarized for step- and minute-based goals using separate models for the graded PA goal phase (weeks 3-8), fixed PA goal phase (weeks 9-24), and across all time points (weeks 3-24). Models were conditioned on the probability of maintaining the same adherence status from the preceding week or transitioning to another status within each type of goal prescription.

The likelihood of having an adherent, suboptimally adherent, or nonadherent status over time was examined for step- and minute-based goals, separately, using multinomial logistic regression for correlated data. The odds of having a less adherent status were modeled for each type of PA recommendation during the graded PA goal phase, fixed PA goal phase, and across all time points. Comparisons were then made between models for step- versus minute-based goals on the odds of transitioning into a less adherent status. All analyses were performed using SAS 9.4 (Cary, NC, USA), and the level of statistical significance was set at $P<.05$.

Results

Participant Characteristics

All participants randomized to the control group (N=212) in the larger RCT [17] were included in this study (Table 2). On average, participants in the current analyses were 47.9 years old, with a BMI of 35.8 kg/m^2. The majority (n=194, 91.5%) were female, and approximately one-third (n=67, 31.6%) self-identified as a racial/ethnic minority. Most participants had at least a college degree (n=169, 79.7%) and were employed full-time (n=152, 71.7%). Retention was 81.1% (n=172) at the 6-month assessment visit.

Table 2. Baseline sociodemographic characteristics of study participants (N=212).

Characteristic	Value
Age (years), mean (SD)	47.9 (11.1)
BMI[a] (kg/m^2), mean (SD)	35.8 (5.9)
Gender, n (%)	
Female	194 (91.5)
Male	18 (8.5)
Race/ethnicity, n (%)	
White	145 (68.4)
Minority[b]	67 (31.6)
Education, n (%)	
College degree or higher	169 (79.7)
Some college or less	43 (20.3)
Marital status, n (%)	
Married or cohabiting	118 (55.7)
Separated, divorced, widowed, or never married	94 (44.3)
Employment status, n (%)	
Full time	152 (71.7)
Part time or unemployed	60 (28.3)
Geographic region, n (%)	
Northeast (Vermont)	106 (50.0)
Southeast (South Carolina)	106 (50.0)

[a]BMI: body mass index.

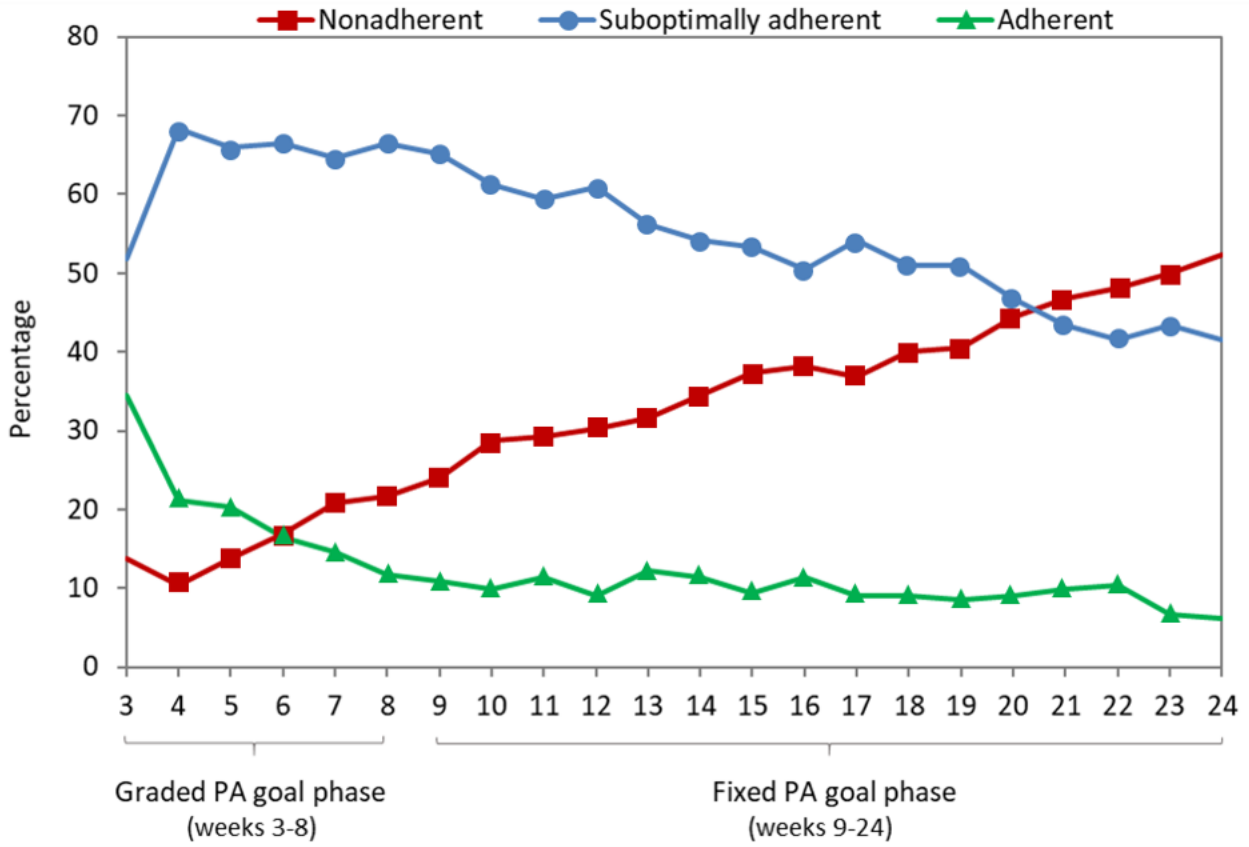
[b]Minority groups include African American, Asian, Hispanic, and Pacific Islander.

Rates of Adherence to Physical Activity

Step-Based Goals

On average, of the 212 participants, 144 (67.7%) self-monitored their steps on ≥1 day/week during the 6-month period and 27 (12.4%) met the prescribed steps goal each week. Approximately one-third (n=73, 34.4%) of the participants started out adherent to the 7000-daily-step goal in week 3 (Figure 1). Rates declined as the goal progressed to 10,000 daily steps, and 13 (6.1%) participants remained adherent by 6 months. More than half of the participants (n=110, 51.9%) initially self-monitored their steps but did not meet the prescribed step goal (suboptimally adherent). The same pattern observed for adherent participants was also seen for suboptimally adherent participants, in which the proportion of suboptimally adherent participants declined over time to 88 (41.5%) participants. In contrast, the proportion of participants who were initially nonadherent (n=29, 13.7%) steadily increased until 111 (52.4%) participants were no longer self-monitoring steps at the conclusion of the 6-month study period.

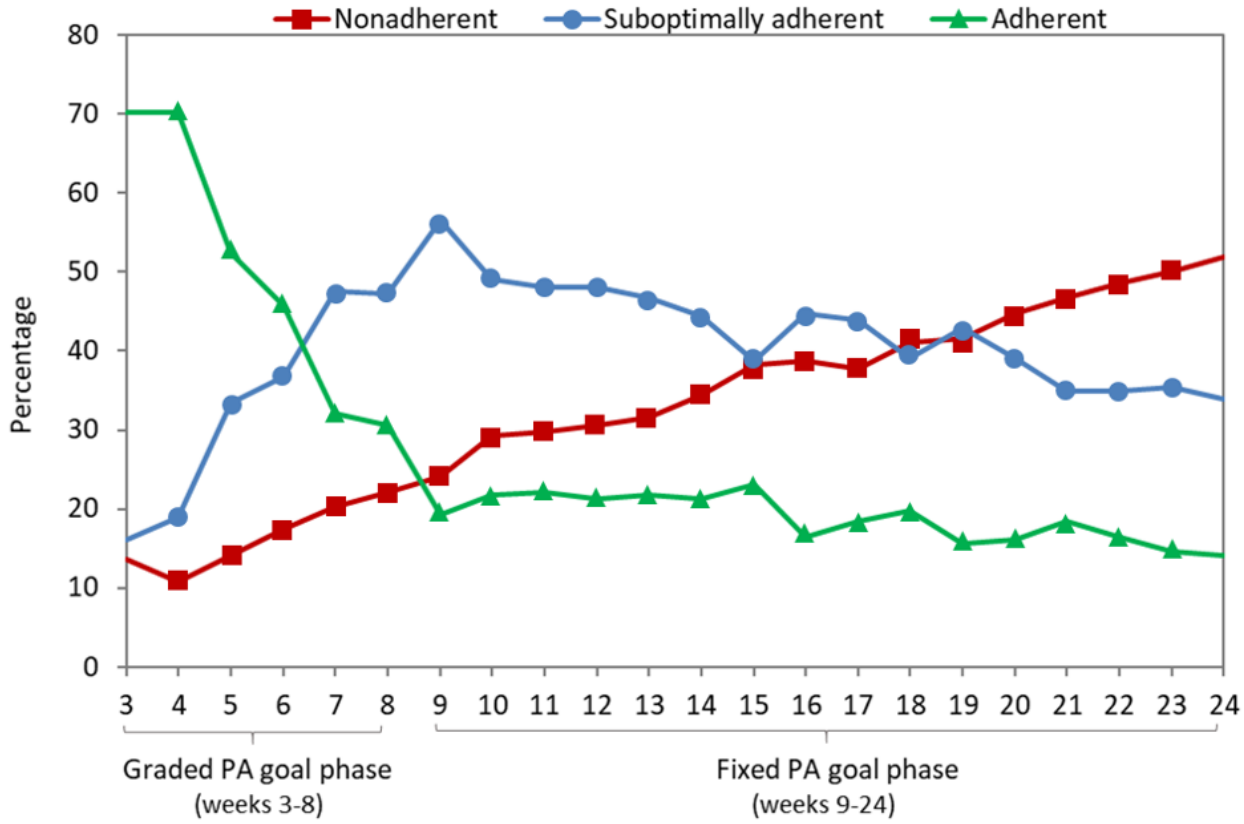
Figure 1. Weekly marginal proportions of participants in each PA adherence category for step-based goals (N=212). PA: physical activity.



Minute-Based Goals

Comparable to daily steps, 143 (67.4%) participants self-monitored their total minutes of MVPA on ≥1 day per week over the 6-month period (Figure 2); however, more than twice as many participants (n=58, 27.4%) met the weekly MVPA minute goals relative to the step goals. Most participants (n=149, 70.3%) were adherent to the initial 50-minute MVPA goal. Adherent behavior rapidly declined as MVPA goals progressed through the graded phase, followed by a slower rate of deterioration in the fixed goal phase to 30 (14.2%) participants by 6 months. The proportion of those classified as suboptimally adherent increased sharply during the graded phase before gradually declining throughout the remaining weeks. Finally, the proportion of participants considered nonadherent to MVPA minute goals closely mirrored that of participants considered nonadherent to step goals over time.

Figure 2. Weekly marginal proportions of participants in each PA adherence category for minute-based goals (N=212). PA: physical activity.



Transitions in Adherence to Physical Activity

The weekly transitional probabilities of adherence to prescribed PA goals and PA self-monitoring across successive weeks of the behavioral weight control program are summarized for steps and minutes of MVPA in Table 3 (see Multimedia Appendix 2 for flow diagrams of transitions). The most stable status by phase and across the full 6-month period for both step- and minute-based goals was the nonadherent status. If a participant was nonadherent to the initial program goals at week 3, they were estimated to remain nonadherent at the end of the 8-week graded phase 90% of the time and transition to suboptimally adherent 10% of the time. The likelihood of remaining nonadherent increased in the fixed goal phase to 98%. A participant had a 100% probability of being nonadherent at week 24, given that they had been nonadherent to the initial PA goals at week 3. It was improbable that a participant would transition directly from a nonadherent status in 1 week to an adherent status the following week.

Table 3. Summary of weekly transitional probabilities of PA[a] adherence by program phase[b].

Adherence status	Nonadherent ($time_t$)		Suboptimally adherent ($time_t$)		Adherent ($time_t$)	
	Steps	MVPA[c]	Steps	MVPA	Steps	MVPA
Graded PA goal phase[d] ($time_{t-1}$), %						
Nonadherent	*89.7* [e]	*89.7*	10.3	10.3	0	0
Suboptimally adherent	15.5	14.7	*80.9*	*70.6*	3.6	14.7
Adherent	4.1	10.7	67.1	49.0	*28.8*	*40.3*
Fixed PA goal phase[f] ($time_{t-1}$), %						
Nonadherent	*97.8*	*97.9*	2.2	0	0	2.1
Suboptimally adherent	44.0	46.0	*52.5*	*48.0*	3.5	6.0
Adherent	16.0	27.7	52.0	36.9	*32.0*	*35.4*
Full 6-month period[g] ($time_{t-1}$), %						
Nonadherent	*100*	*100*	0	0	0	0
Suboptimally adherent	53.6	64.7	*44.5*	*29.4*	1.8	5.9
Adherent	31.5	39.6	53.4	41.6	*15.1*	*18.8*

[a]PA: physical activity.

[b]Transition probabilities describe the likelihood of remaining in an adherence status in 1 week ($time_t$), given the adherence status in the previous week ($time_{t-1}$) or transitioning to another status.

[c]MVPA: moderate-to-vigorous physical activity.

[d]Probabilities indicate adherence status at the end of the graded PA goal phase (week 8), given the adherence status for the initial PA goal at week 3.

[e]Values in italics indicate remaining in the same adherence status.

[f]Probabilities indicate adherence status at the end of the fixed PA goal phase (week 24), given the adherence status at the end of the graded PA goal phase (week 8).

[g]Probabilities indicate adherence status at the end of the study period (week 24), given the adherence status for the initial PA goal at week 3.

Step-Based Goals

During the initial 8 weeks, as the goals progressed from 7000 to 10,000 daily steps, the probability that a participant would remain adherent was 29%, while the probability of transitioning from adherent to suboptimally adherent was 67% (Table 3; also see Multimedia Appendix 2A for step-based goals). For example, if a participant was suboptimally adherent to the 7000-step goal at week 3, they had an 81% chance of remaining suboptimally adherent and a 16% chance of transitioning to a nonadherent status when the goal increased to 10,000 steps at week 8. Across the fixed goal phase, those who were suboptimally adherent at the end of the graded phase (week 8) had a 53% chance of continuing to be suboptimally adherent, a 44% chance of transitioning to a nonadherent status, and a 4% chance of transitioning to an adherent status at the conclusion of 6 months. A participant was expected to remain adherent at the end of the program 15% of the time if they were adherent in week 3; however, the chances of being adherent at the end of the program increased to 32% if they were adherent at week 8. When looking at weekly transitions during the phase with graded PA goals, participants had the highest likelihood of remaining suboptimally adherent or transitioning from being nonadherent to suboptimally adherent between weeks 3 and 4 (95% and 31%, respectively), as shown in Table 4 (see Multimedia Appendix 3A for step-based goals).

Table 4. Weekly transitional probabilities of PA[a] adherence during the graded goal phase[b].

Adherence status	Nonadherent ($time_t$)		Suboptimally adherent ($time_t$)		Adherent ($time_t$)	
	Steps	MVPA[c]	Steps	MVPA	Steps	MVPA
Week 4 ($time_{t-1}$), %						
Nonadherent	*69.0* [d]	*72.4*	31.0	24.1	0	3.4
Suboptimally adherent	1.8	29.0	*94.5*	*50.0*	3.6	47.1
Adherent	0	0.7	43.8	10.7	*56.2*	*88.6*
Week 5 ($time_{t-1}$), %						
Nonadherent	*81.8*	*82.6*	18.2	17.4	0	0
Suboptimally adherent	7.6	22.5	*82.1*	*62.5*	10.3	15.0
Adherent	0	1.3	37.8	28.2	*62.2*	*70.5*
Week 6 ($time_{t-1}$), %						
Nonadherent	*93.1*	*93.3*	6.9	6.7	0	0
Suboptimally adherent	5.7	9.9	*86.4*	*69.0*	7.9	21.1
Adherent	2.3	1.8	41.9	24.3	*55.8*	*73.9*
Week 7 ($time_{t-1}$), %						
Nonadherent	*86.1*	*83.8*	13.9	13.5	0	2.7
Suboptimally adherent	9.2	14.1	*82.3*	*78.2*	8.5	7.7
Adherent	0	1.0	45.7	36.1	*54.3*	*62.9*
Week 8 ($time_{t-1}$), %						
Nonadherent	*81.8*	*81.4*	18.2	18.6	0	0
Suboptimally adherent	7.3	10.9	*85.4*	*71.3*	7.3	17.8
Adherent	0	1.5	51.6	29.4	*48.4*	*69.1*

[a]PA: physical activity.

[b]Transition probabilities describe the likelihood of remaining in an adherence status in 1 week ($time_t$), given the adherence status in the previous week ($time_{t-1}$) or transitioning to another status.

[c]MVPA: moderate-to-vigorous physical activity.

[d]Values in italics indicate remaining in the same adherence status.

Minute-Based Goals

In the graded goal phase, participants had a 40% chance of remaining adherent to the 150-minute recommendation at week 8 if they were initially adherent to the 50-minute target, in addition to a 49% chance of transitioning from being adherent to suboptimally adherent (Table 3; also see Multimedia Appendix 2B for minute-based goals). Participants were likely to remain suboptimally adherent at week 8 approximately 71% of the time if their initial status had been suboptimally adherent, with an equal chance of moving from a suboptimally adherent to a nonadherent or adherent status. During the period when the prescribed MVPA was fixed at 200 minutes per week (weeks 9-24), a participant was most likely to remain suboptimally adherent (48%) or transition from being suboptimally adherent to nonadherent (46%), with a 6% chance of transitioning to an adherent status. A participant was expected to remain adherent at the end of the program 19% of the time, given that they were adherent to the initial 50-minute MVPA goal, and doubled their chances of being adherent at the end of the program if they were adherent at the 8-week mark. When looking at weekly transitions during the phase with graded PA goals, adherence status from weeks 3 to 4 demonstrated the greatest chances of remaining adherent (89%), improving from suboptimally adherent to adherent (47%), and improving from nonadherent to suboptimally adherent (24%). See Table 4; in addition, see Multimedia Appendix 3B for minute-based goals.

Comparison of Adherence Categories for Step-Based Versus Minute-Based Goals

Analyses examining adherence status over time indicated that participants were significantly more likely to be in a less adherent category relative to adherent as the program progressed (Table 5). Thus, participants were most likely to transition to a lower adherence category each week in both goal phases and across the 6-month period for step-based goals (all $P<.001$), as well as for minute-based goals (all $P<.001$).

Table 5. Comparison of step- versus minute-based goals on the odds of being in a less adherent weekly status.

Phase	Step-based goals	Minute-based goals	Steps vs minutes
	OR[a] (95% CI)	OR (95% CI)	*P* value[b]
Graded PA[c] goal phase[d]	*1.24 (1.17-1.30)* [e]	*1.39 (1.31-1.48)*	.004
Fixed PA goal phase[f]	*1.07 (1.05-1.08)*	*1.06 (1.05-1.08)*	.65
Full 6-month period[g]	*1.09 (1.09-1.09)*	*1.11 (1.11-1.11)*	.07

[a]OR: odds ratio (odds of being in a less adherent category over time; "adherent" is the reference group).

[b]*P* value for the odds of being in a less adherent category, relative to "adherent," during the specified phase for step-based goals versus minute-based goals.

[c]PA: physical activity.

[d]Graded PA goal phase includes weeks 3-8.

[e]Values in italics indicate statistical significance at $P<.001$ during the specified phase for each type of goal.

[f]Fixed PA goal phase includes weeks 9-24.

[g]Full 6-month period includes weeks 3-24.

For steps, participants were 1.24 times more likely to be less adherent, relative to being adherent, each week during the graded goal phase, 1.07 times more likely to shift to a less adherent status during the fixed goal phase, and 1.09 times more likely to be in a less adherent category each week across the 6-month period. For minutes of MVPA, participants were 1.39 and 1.06 times more likely to be in a less adherent category each week during the graded and fixed goal phases, respectively, and 1.11 times more likely to shift to a less adherent category each week over the entire 6 months.

When steps and minutes of MVPA were compared during the graded PA goal phase, participants were significantly more likely to regress to a less adherent category the following week when examining minutes of MVPA relative to considering steps ($P=.004$). In other words, participants were more likely to become less adherent from week to week with respect to minute-based goals compared to step-based goals during weeks 3-8. No significant difference was seen in the propensity to be adherent between step- and minute-based goals during the fixed goal phase ($P=.65$). Although a trend was noted between models for steps and minutes of MVPA in the odds of being in a less adherent category when examining the full 6-month period ($P=.07$), the difference was not statistically significant.

Discussion

Principal Findings

To the best of our knowledge, this study is the first to describe the stability of adherence to weekly self-monitoring and attainment of PA goals prescribed in terms of daily steps and weekly minutes of MVPA during an online behavioral weight control program. Few participants met the initial step-based goals and subsequently either remained in a nonadherent or a suboptimally adherent status from one week to the next or transitioned from a suboptimally adherent to a nonadherent status over time. Conversely, most participants started out adherent to the minute-based goals but demonstrated greater variability between adherence statuses from week to week than was seen for step-based goals. Thus, participants with overweight and obesity enrolled in a behavioral weight control program were unlikely to achieve even the lowest step-based goals or to catch up with the goals as the step target progressively increased but had a greater chance of meeting the minute-based goals, particularly early in treatment.

The weekly transitional probabilities identified in this study underscore the initial 2 months of a lifestyle program as a formative period that contributes to an individual's chances of future PA adherence. Participants who met the earliest recommendations were most likely to be adherent at 6 months relative to those who started out suboptimally adherent or nonadherent. Intriguingly, the likelihood of long-term adherence increased 2-fold if a participant remained adherent through the graded goal phase. Our findings substantiate previous studies that demonstrate that the first 2 months of program initiation serve as a critical juncture associated with better long-term program adherence, weight loss, and health-related outcomes [16,27-30]. From a theoretical perspective, successful adoption of PA recommendations may be attributed to an increase in self-efficacy and mastery experiences during program initiation from frequent, intentional exposure to self-monitoring and goal attainment [31-33]. Future studies should examine whether targeting this early period with strategies to facilitate consistent self-monitoring in conjunction with attaining recommendations substantially increases a person's chances of long-term success in achieving PA targets and in losing weight.

Few participants achieved stability in adherent behavior, which is of concern. Indeed, nonadherence was the most consistent status and was increasingly more likely over time, regardless of the type of PA goal. Some adults with overweight may prefer monitoring minute-based activity rather than step counts [34]; however, our study did not detect a difference in self-monitoring behavior between these 2 types of PA metrics. It may be that the method of submitting daily step totals and minutes of MVPA concurrently to the study website's digital diary contributed to the matching rates of self-monitoring steps and minutes of MVPA. Nevertheless, a considerable proportion of participants failed to track their activity and were unlikely to reengage with self-monitoring despite regular reminders of program PA goals and prompts to self-monitor, which were given in weekly emailed feedback, group meetings, and online lessons. Providing

wearable devices to those who fail to engage in manual self-monitoring methods may improve adherence by facilitating continuous PA tracking and minimizing the burden associated with manual self-monitoring [35,36]. However, others have demonstrated that providing a wearable activity-tracking device in an online weight control program does not promote greater PA engagement or weight loss [37].

Suboptimally adherent behavior appeared to be the most vulnerable to transitions, particularly with respect to minute-based goals. It is striking that participants had the best chances of shifting from suboptimally adherent to fully adherent behavior during the graded goal phase, yet the likelihood of becoming nonadherent outweighed becoming adherent after 2 months. A similar pattern with weekly dietary self-monitoring was previously reported where individuals who demonstrated suboptimal adherence to dietary self-monitoring had the greatest potential for transitioning to adherent behavior during the first 2 months of a lifestyle program but not after that time [16]. It is possible that a vulnerability to transitions during program initiation represents an intervention opportunity to increase the likelihood of shifting to fully adherent behavior. Supporting individuals with coaching sessions at the first sign of vulnerability might foster an increase in their MVPA or step counts [38-40]. A change in status from adherent to suboptimally adherent exemplifies 1 event that triggers implementation of a coaching session to quickly bring the individual back on a trajectory of success and avoid a decline to nonadherent behavior when it may be too late to reengage them in program recommendations.

To capitalize on the early weeks of program initiation and move more individuals toward adherent behavior, reevaluating how PA recommendations are structured for behavioral weight control is warranted. It is apparent that providing everyone with the same incremental, linear progression of PA goals does not accommodate the dynamic process of lifestyle change for each individual. From a clinical perspective, it is imperative to appreciate that the continuation of conventional treatment likely will not help vulnerable subgroups transition to adherent behavior. Instead, it may be important to provide tailored goals that adapt based on the individual's recent PA, thus providing more realistic, attainable PA targets [38,41-43].

Although this study took the formative step of describing behavioral patterns, more work is needed to identify precisely how best to adapt treatment to capitalize on these patterns and enhance outcomes. There are likely other factors, such as cognitive, affective, and motivational components, influencing whether individuals are adherent to PA, and these should be explored to guide treatment tailoring [44-48]. For example, does exercise become unpleasant or untenable for some individuals when MVPA goals progress to 150 minutes, which contributes to them no longer adhering to recommendations? Do other individuals experience boredom with their exercise routine or a lack of motivation to pursue the goals? To what extent does a sense of self-efficacy impact attainment of step- or minute-based goals? There may also be other contextual factors influencing adherence, such as a vacation, illness/injury, or other barriers to PA, which interrupt a person's routine. Furthermore, there may be dynamic bidirectional influences of weight loss success and PA goal attainment in the context of a weight control program; examination of these dynamic processes is warranted because success with weight loss may well be a driver of PA goal attainment, in addition to PA goal attainment driving weight loss success [1,9,11]. More research is needed before firm clinical conclusions can be made about how to adjust PA recommendations in behavioral weight control.

Limitations

This research advances our understanding of the effects of different types of PA goals commonly prescribed in behavioral weight control interventions on goal attainment combined with self-monitoring among a large sample of adults with overweight and obesity. However, findings should be considered with the following limitations in mind. First, participants were predominantly women with obesity, and results may not generalize to other populations. Next, the influence of exercise history or the level of activity when entering the study on adherence patterns could not be determined. If program goal attainment differs between individuals who are inactive at study entry compared to those who are already engaging in some PA at baseline, recommendations may be tailored based on baseline activity levels. Although the goals offered in the program are similar to those provided by other behavioral weight control programs [5,7], patterns of goal attainment noted in this study may not generalize to other programs that recommend substantively different types or doses of PA or to PA promotion programs without an emphasis on weight loss. It is possible that the MVPA goal was more achievable than the step goal or that the participants preferred MVPA or preferred tracking MVPA. Additionally, other programs have used different MVPA goals, and it will be important to determine whether these findings generalize to other, usually higher, MVPA goals. Exploring the behavioral patterns of adherence to other PA recommendations and programs will be advantageous in determining whether there are specific thresholds for the number of steps and minutes of MVPA that optimize adherence. In addition, this study provided both PA goals, and it will be important in future research to determine if one goal or the other is more effective in inducing weight loss. Furthermore, it is possible that some participants categorized as nonadherent in a given week did, in fact, meet the corresponding PA goal for that week; however, goal attainment could not be determined in this case if they did not submit their activity record on the website. Further, goal attainment relied on self-report of PA, which is well known for overestimation [49,50]. Providing technology in future studies to assist with objective PA measurement and self-monitoring could remove some of the barriers associated with manual posting of self-reported PA. Finally, a participant's adherence status for each type of PA goal prescription was conditional on the preceding week's level of adherence for that prescription and calculated independently from adherence to the other type of goal prescription; the influence of adherence at other time points or adherence to the other PA goal is an area for future study. Nevertheless, the approach selected allows for a preliminary examination of dynamic adherence to PA goals and self-monitoring.

Conclusion

This study begins to identify key transition points in the behavior change process as it relates to PA adherence for step- and minute-based goals in a behavioral weight loss program and provides a useful framework for detecting responsiveness to program recommendations. Moreover, it may assist in shaping more effective PA recommendations in lifestyle programs, particularly during the early period of program initiation. The initial 2 months provide a window of opportunity to assist people in transitioning to adherent behavior, especially those who fail to meet the PA goals but engage in self-monitoring. As a result, researchers and clinicians should consider PA self-monitoring and goal attainment concurrently as an indication of an individual's adherence status. Emphasis should be placed on strategies that help people start and reach the 2-month mark with consistent self-monitoring and attainment of PA recommendations, since this appears to be a formative period for continued adherence. Future research should investigate how these transitional patterns in PA adherence status are influenced by individual characteristics and contextual factors, as well as how they relate to weight loss and other health-related outcomes, to inform optimization of treatment recommendations.

Acknowledgments

This research was supported in part by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK; grant no R01DK056746) to JH and DW. The NIDDK had no role in the study design; collection, analysis, and interpretation of data; writing of the paper; or the decision to submit for publication.

Conflicts of Interest

RK reports personal fees and nonfinancial support from General Mills unrelated to the submitted work. All other authors declare no conflict of interests.

Multimedia Appendix 1
CONSORT diagram.
[PDF File (Adobe PDF File), 83 KB - jmir_v24i1e30673_app1.pdf]

Multimedia Appendix 2
Flow diagram summarizing transitions in weekly adherence status to (A) step-based goals and (B) minute-based goals by physical activity (PA) goal phase (N=212). Solid bars represent the marginal proportions of participants in each adherence category at a given time point; bands represent the proportion transitioning to another category or remaining in the same category at 1 time point, given the previous time point.
[PNG File , 500 KB - jmir_v24i1e30673_app2.png]

Multimedia Appendix 3
Flow diagram of transitions in weekly adherence status to (A) step-based goals and (B) minute-based goals across the graded physical activity (PA) goal phase (N=212). Solid bars represent the marginal proportions of participants in each adherence category at a given time point; bands represent the proportion transitioning to another category or remaining in the same category at 1 time point, given the previous time point.
[PNG File , 872 KB - jmir_v24i1e30673_app3.png]

References

1. Wing RR, Hamman RF, Bray GA, Delahanty L, Edelstein SL, Hill JO, Diabetes Prevention Program Research Group. Achieving weight and activity goals among diabetes prevention program lifestyle participants. Obes Res 2004 Sep;12(9):1426-1434 [FREE Full text] [doi: 10.1038/oby.2004.179] [Medline: 15483207]
2. Acharya SD, Elci OU, Sereika SM, Music E, Styn MA, Turk MW, et al. Adherence to a behavioral weight loss treatment program enhances weight loss and improvements in biomarkers. Patient Prefer Adherence 2009 Nov 03;3:151-160 [FREE Full text] [doi: 10.2147/ppa.s5802] [Medline: 19936157]
3. 2018 Physical Activity Guidelines Advisory Committee. 2018 Physical Activity Guidelines Advisory Committee Scientific Report. Washington, DC: U.S. Department of Health and Human Services; 2018. URL: https://health.gov/sites/default/files/2019-09/PAG_Advisory_Committee_Report.pdf [accessed 2021-06-06]
4. Jensen MD, Ryan DH, Apovian CM, Ard JD, Comuzzie AG, Donato KA, American College of Cardiology/American Heart Association Task Force on Practice Guidelines, Obesity Society. 2013 AHA/ACC/TOS guideline for the management of overweight and obesity in adults: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and The Obesity Society. Circulation 2014 Jun 24;129(25 Suppl 2):S102-S138 [FREE Full text] [doi: 10.1161/01.cir.0000437739.71477.ee] [Medline: 24222017]

5. Look AHEAD Research Group, Wadden TA, West DS, Delahanty L, Jakicic J, Rejeski J, et al. The Look AHEAD study: a description of the lifestyle intervention and the evidence supporting it. Obesity (Silver Spring) 2006 May;14(5):737-752 [FREE Full text] [doi: 10.1038/oby.2006.84] [Medline: 16855180]
6. Butryn ML, Webb V, Wadden TA. Behavioral treatment of obesity. Psychiatr Clin North Am 2011 Dec;34(4):841-859 [FREE Full text] [doi: 10.1016/j.psc.2011.08.006] [Medline: 22098808]
7. Diabetes Prevention Program (DPP) Research Group. The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care 2002 Dec;25(12):2165-2171 [FREE Full text] [doi: 10.2337/diacare.25.12.2165] [Medline: 12453955]
8. U.S. Department of Health and Human Services. Physical Activity Guidelines for Americans, 2nd edition. 2018. URL: https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf [accessed 2021-06-06]
9. Wadden T, West D, Neiberg R, Wing RR, Ryan DH, Johnson KC, Look AHEAD Research Group. One-year weight losses in the Look AHEAD study: factors associated with success. Obesity (Silver Spring) 2009 Apr;17(4):713-722 [FREE Full text] [doi: 10.1038/oby.2008.637] [Medline: 19180071]
10. Conroy M, Yang K, Elci O, Gabriel KP, Styn MA, Wang J, et al. Physical activity self-monitoring and weight loss: 6-month results of the SMART trial. Med Sci Sports Exerc 2011 Aug;43(8):1568-1574 [FREE Full text] [doi: 10.1249/MSS.0b013e31820b9395] [Medline: 21200337]
11. Ritchie ND, Carroll JK, Holtrop JS, Havranek EP. Effects of physical activity goal attainment on engagement and outcomes in the National Diabetes Prevention Program. Transl Behav Med 2018 Nov 21;8(6):932-937. [doi: 10.1093/tbm/ibx021] [Medline: 29669050]
12. Unick J, Gaussoin S, Hill J, Jakicic JM, Bond DS, Hellgren M, et al. Four-year physical activity levels among intervention participants with type 2 diabetes. Med Sci Sports Exerc 2016 Dec;48(12):2437-2445 [FREE Full text] [doi: 10.1249/MSS.0000000000001054] [Medline: 27471785]
13. Wing RR. Behavioral approaches to the treatment of obesity. In: Bray GA, Bouchard C, editors. Handbook of Obesity: Clinical Applications. 2nd ed. New York, NY: Guilford Press; 2004:147-167.
14. Burke LE, Wang J, Sevick MA. Self-monitoring in weight loss: a systematic review of the literature. J Am Diet Assoc 2011 Jan;111(1):92-102 [FREE Full text] [doi: 10.1016/j.jada.2010.10.008] [Medline: 21185970]
15. Gavin KL, Sherwood NE, Wolfson J, Pereira MA, Linde JA. Characterizing self-monitoring behavior and its association with physical activity and weight loss maintenance. Am J Lifestyle Med 2021;15(2):173-183 [FREE Full text] [doi: 10.1177/1559827618790556] [Medline: 33786033]
16. Sevick MA, Stone RA, Zickmund S, Wang Y, Korytkowski M, Burke LE. Factors associated with probability of personal digital assistant-based dietary self-monitoring in those with type 2 diabetes. J Behav Med 2010 Aug;33(4):315-325. [doi: 10.1007/s10865-010-9257-9] [Medline: 20232131]
17. West DS, Krukowski RA, Finkelstein EA, Stansbury ML, Ogden DE, Monroe CM, et al. Adding financial incentives to online group-based behavioral weight control: an RCT. Am J Prev Med 2020 Aug;59(2):237-246 [FREE Full text] [doi: 10.1016/j.amepre.2020.03.015] [Medline: 32446752]
18. Harvey-Berino J, West D, Krukowski R, Prewitt E, VanBiervliet A, Ashikaga T, et al. Internet delivered behavioral obesity treatment. Prev Med 2010 Aug;51(2):123-128 [FREE Full text] [doi: 10.1016/j.ypmed.2010.04.018] [Medline: 20478333]
19. West DS, Harvey JR, Krukowski RA, Prewitt TE, Priest J, Ashikaga T. Do individual, online motivational interviewing chat sessions enhance weight loss in a group-based, online weight control program? Obesity (Silver Spring) 2016 Nov;24(11):2334-2340 [FREE Full text] [doi: 10.1002/oby.21645] [Medline: 27616628]
20. Tudor-Locke C, Craig CL, Brown WJ, Clemes SA, De Cocker K, Giles-Corti B, et al. How many steps/day are enough? For adults. Int J Behav Nutr Phys Act 2011 Jul 28;8:79 [FREE Full text] [doi: 10.1186/1479-5868-8-79] [Medline: 21798015]
21. Downs D, LeMasurier G, DiNallo J. Baby steps: pedometer-determined and self-reported leisure-time exercise behaviors of pregnant women. J Phys Act Health 2009 Jan;6(1):63-72. [doi: 10.1123/jpah.6.1.63] [Medline: 19211959]
22. Rowe DA, Mahar MT, Raedeke TD, Lore J. Measuring physical activity in children with pedometers: reliability, reactivity, and replacement of missing data. Pediatr Exerc Sci 2004 Nov;16(4):343-354. [doi: 10.1123/pes.16.4.343]
23. Slagter SN, Corpeleijn E, van der Klauw MM, Sijtsma A, Swart-Busscher LG, Perenboom CWM, et al. Dietary patterns and physical activity in the metabolically (un)healthy obese: the Dutch Lifelines cohort study. Nutr J 2018 Feb 12;17(1):18 [FREE Full text] [doi: 10.1186/s12937-018-0319-0] [Medline: 29433580]
24. Centers for Disease Control and Prevention. A Data Users Guide to the BRFSS Physical Activity Questions: How to Assess the 2008 Physical Activity Guidelines for Americans. 2013. URL: https://www.cdc.gov/brfss/pdf/PA%20RotatingCore_BRFSSGuide_508Comp_07252013FINAL.pdf [accessed 2021-06-06]
25. Wang J, Sereika SM, Chasens ER, Ewing LJ, Matthews JT, Burke LE. Effect of adherence to self-monitoring of diet and physical activity on weight loss in a technology-supported behavioral intervention. Patient Prefer Adherence 2012;6:221-226 [FREE Full text] [doi: 10.2147/PPA.S28889] [Medline: 22536058]
26. Tudor-Locke C, Craig CL, Thyfault JP, Spence JC. Appl Physiol Nutr Metab 2013 Feb;38(2):100-114 [FREE Full text] [doi: 10.1139/apnm-2012-0235] [Medline: 23438219]
27. Unick JL, Ross KM, Wing RR. Factors associated with early non-response within an internet-based behavioural weight loss program. Obes Sci Pract 2019 Aug;5(4):324-332 [FREE Full text] [doi: 10.1002/osp4.341] [Medline: 31452917]

28. Unick JL, Pellegrini CA, Demos KE, Dorfman L. Initial weight loss response as an indicator for providing early rescue efforts to improve long-term treatment outcomes. Curr Diab Rep 2017 Sep;17(9):69 [FREE Full text] [doi: 10.1007/s11892-017-0904-1] [Medline: 28726155]
29. Unick JL, Neiberg RH, Hogan PE, Cheskin LJ, Dutton GR, Jeffery R, Look AHEAD Research Group. Weight change in the first 2 months of a lifestyle intervention predicts weight changes 8 years later. Obesity (Silver Spring) 2015 Jul;23(7):1353-1356 [FREE Full text] [doi: 10.1002/oby.21112] [Medline: 26110890]
30. Unick JL, Walkup MP, Miller ME, Apolzan JW, Brubaker PH, Coday M, Look AHEAD Research Group. Early physical activity adoption predicts longer-term physical activity among individuals inactive at baseline. J Phys Act Health 2020 Nov 03;17(12):1205-1212. [doi: 10.1123/jpah.2020-0114] [Medline: 33152692]
31. Latham GP, Locke EA. Self-regulation through goal setting. Organ Behav Hum Decis Proc 1991 Dec;50(2):212-247. [doi: 10.1016/0749-5978(91)90021-K]
32. Locke EA, Latham GP. New directions in goal-setting theory. Curr Dir Psychol Sci 2016 Jun 23;15(5):265-268. [doi: 10.1111/j.1467-8721.2006.00449.x]
33. Strecher VJ, Seijts GH, Kok GJ, Latham GP, Glasgow R, DeVellis B, et al. Goal setting as a strategy for health behavior change. Health Educ Q 1995 May;22(2):190-200. [doi: 10.1177/109019819502200207] [Medline: 7622387]
34. Fu E, White MA, Hughto JMW, Steiner B, Willis EA. Development of the Physical Activity Tracking Preference Questionnaire. Int J Exerc Sci 2019;12(5):297-309 [FREE Full text] [Medline: 31156755]
35. Butryn ML, Godfrey KM, Martinelli MK, Roberts SR, Forman EM, Zhang F. Digital self-monitoring: does adherence or association with outcomes differ by self-monitoring target? Obes Sci Pract 2020 Apr;6(2):126-133 [FREE Full text] [doi: 10.1002/osp4.391] [Medline: 32313670]
36. Carels RA, Darby LA, Rydin S, Douglass OM, Cacciapaglia HM, O'Brien WH. The relationship between self-monitoring, outcome expectancies, difficulties with eating and exercise, and physical activity and weight loss treatment outcomes. Ann Behav Med 2005 Dec;30(3):182-190. [doi: 10.1207/s15324796abm3003_2] [Medline: 16336069]
37. Thomas JG, Raynor HA, Bond DS, Luke AK, Cardoso CC, Foster GD, et al. Weight loss in Weight Watchers Online with and without an activity tracking device compared to control: A randomized trial. Obesity (Silver Spring) 2017 Jun 24;25(6):1014-1021. [doi: 10.1002/oby.21846] [Medline: 28437597]
38. Pfammatter AF, Nahum-Shani I, DeZelar M, Scanlan L, McFadden HG, Siddique J, et al. SMART: study protocol for a sequential multiple assignment randomized controlled trial to optimize weight loss management. Contemp Clin Trials 2019 Jul;82:36-45 [FREE Full text] [doi: 10.1016/j.cct.2019.05.007] [Medline: 31129369]
39. Fischer X, Kreppke JN, Zahner L, Gerber M, Faude O, Donath L. Telephone-based coaching and prompting for physical activity: short- and long-term findings of a randomized controlled trial (Movingcall). Int J Environ Res Public Health 2019 Jul 23;16(14):2626 [FREE Full text] [doi: 10.3390/ijerph16142626] [Medline: 31340528]
40. Alley S, Jennings C, Plotnikoff RC, Vandelanotte C. Web-based video-coaching to assist an automated computer-tailored physical activity intervention for inactive adults: a randomized controlled trial. J Med Internet Res 2016 Aug 12;18(8):e223 [FREE Full text] [doi: 10.2196/jmir.5664] [Medline: 27520283]
41. Korinek EV, Phatak SS, Martin CA, Freigoun MT, Rivera DE, Adams MA, et al. Adaptive step goals and rewards: a longitudinal growth model of daily steps for a smartphone-based walking intervention. J Behav Med 2018 Feb;41(1):74-86. [doi: 10.1007/s10865-017-9878-3] [Medline: 28918547]
42. Poirier J, Bennett WL, Jerome GJ, Shah NG, Lazo M, Yeh H, et al. Effectiveness of an activity tracker- and internet-based adaptive walking program for adults: a randomized controlled trial. J Med Internet Res 2016 Feb 09;18(2):e34 [FREE Full text] [doi: 10.2196/jmir.5295] [Medline: 26860434]
43. Adams MA, Sallis JF, Norman GJ, Hovell MF, Hekler EB, Perata E. An adaptive physical activity intervention for overweight adults: a randomized controlled trial. PLoS One 2013 Dec;8(12):e82901 [FREE Full text] [doi: 10.1371/journal.pone.0082901] [Medline: 24349392]
44. Dunton GF, Leventhal AM, Rothman AJ, Intille SS. Affective response during physical activity: within-subject differences across phases of behavior change. Health Psychol 2018 Oct;37(10):915-923 [FREE Full text] [doi: 10.1037/hea0000644] [Medline: 30234350]
45. Ekkekakis P, Lind E, Vazou S. Affective responses to increasing levels of exercise intensity in normal-weight, overweight, and obese middle-aged women. Obesity (Silver Spring) 2010 Jan;18(1):79-85 [FREE Full text] [doi: 10.1038/oby.2009.204] [Medline: 19556979]
46. Focht BC. Affective responses to 10-minute and 30-minute walks in sedentary, overweight women: relationships with theory-based correlates of walking for exercise. Psychol Sport Exerc 2013 Sep;14(5):759-766. [doi: 10.1016/j.psychsport.2013.04.003]
47. Unick JL, Michael JC, Jakicic JM. Affective responses to exercise in overweight women: initial insight and possible influence on energy intake. Psychol Sport Exerc 2012 Sep 01;13(5):528-532 [FREE Full text] [doi: 10.1016/j.psychsport.2012.02.012] [Medline: 24039545]
48. Ajzen I, Kruglanski AW. Reasoned action in the service of goal pursuit. Psychol Rev 2019 Oct;126(5):774-786. [doi: 10.1037/rev0000155] [Medline: 31343213]

49. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act 2008 Nov 06;5:56 [FREE Full text] [doi: 10.1186/1479-5868-5-56] [Medline: 18990237]
50. Lichtman SW, Pisarska K, Berman ER, Pestone M, Dowling H, Offenbacher E, et al. Discrepancy between self-reported and actual caloric intake and exercise in obese subjects. N Engl J Med 1992 Dec 31;327(27):1893-1898. [doi: 10.1056/NEJM199212313272701] [Medline: 1454084]

Abbreviations

BMI: body mass index
MVPA: moderate-to-vigorous physical activity
OR: odds ratio
PA: physical activity
RCT: randomized controlled trial

Edited by A Mavragani; submitted 24.05.21; peer-reviewed by Anonymous; comments to author 15.06.21; revised version received 17.06.21; accepted 22.12.21; published 28.01.22.

Please cite as:
Stansbury ML, Harvey J, Krukowski RA, Pellegrini CA, Wang X, West DS
Describing Transitions in Adherence to Physical Activity Self-monitoring and Goal Attainment in an Online Behavioral Weight Loss Program: Secondary Analysis of a Randomized Controlled Trial
J Med Internet Res 2022;24(1):e30673
URL: https://www.jmir.org/2022/1/e30673
doi:10.2196/30673
PMID:35089159

Original Paper

The Effect of Internet-Delivered Cognitive Behavioral Therapy Versus Psychoeducation Only on Psychological Distress in Patients With Noncardiac Chest Pain: Randomized Controlled Trial

Ghassan Mourad[1], RN, PhD; Magda Eriksson-Liebon[1,2], RN, MSc; Patric Karlström[1,3], MD, PhD; Peter Johansson[1,4], RN, PhD

[1]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

[2]Department of Emergency Medicine in Norrköping, and Department of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

[3]Department of Internal Medicine, Ryhov County Hospital, Region Jönköping County, Jönköping, Sweden

[4]Department of Internal Medicine in Norrköping, and Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Ghassan Mourad, RN, PhD
Department of Health, Medicine and Caring Sciences
Linköping University
Linköping
Sweden
Phone: 46 +46736209582
Email: ghassan.mourad@liu.se

Abstract

Background: Patients with recurrent episodes of noncardiac chest pain (NCCP) experience cardiac anxiety as they misinterpret the pain to be cardiac related and avoid physical activity that they think could threaten their lives. Psychological interventions, such as internet-delivered cognitive behavioral therapy (iCBT), targeting anxiety can be a feasible solution by supporting patients to learn how to perceive and handle their chest pain.

Objective: This study aims to evaluate the effects of a nurse-led iCBT program on cardiac anxiety and other patient-reported outcomes in patients with NCCP.

Methods: Patients with at least two health care consultations because of NCCP during the past 6 months, and who were experiencing cardiac anxiety (Cardiac Anxiety Questionnaire score ≥24), were randomized into 5 weeks of iCBT (n=54) or psychoeducation (n=55). Patients were aged 54 (SD 17) years versus 57 (SD 16) years and were mainly women (32/54, 59% vs 35/55, 64%). The iCBT program comprised psychoeducation, mindfulness, and exposure to physical activity, with weekly homework assignments. The primary outcome was cardiac anxiety. The secondary outcomes were fear of bodily sensations, depressive symptoms, health-related quality of life, and chest pain frequency. Intention-to-treat analysis was applied, and the patients were followed up for 3 months. Mixed model analysis was used to determine between-group differences in primary and secondary outcomes.

Results: No significant differences were found between the iCBT and psychoeducation groups regarding cardiac anxiety or any of the secondary outcomes in terms of the interaction effect of time and group over the 3-month follow-up. iCBT demonstrated a small effect size on cardiac anxiety (Cohen *d*=0.31). In the iCBT group, 36% (16/44) of patients reported a positive reliable change score (≥11 points on the Cardiac Anxiety Questionnaire), and thus an improvement in cardiac anxiety, compared with 27% of (13/48) patients in the psychoeducation group. Within-group analysis showed further significant improvement in cardiac anxiety (*P*=.04) at the 3-month follow-up compared with the 5-week follow-up in the iCBT group but not in the psychoeducation group.

Conclusions: iCBT was not superior to psychoeducation in decreasing cardiac anxiety in patients with NCCP. However, iCBT tends to have better long-term effects on psychological distress, including cardiac anxiety, health-related quality of life, and NCCP frequency than psychoeducation. The effects need to be followed up to draw more reliable conclusions.

Trial Registration: ClinicalTrials.gov NCT03336112; https://www.clinicaltrials.gov/ct2/show/NCT03336112

(J Med Internet Res 2022;24(1):e31674) doi:10.2196/31674

KEYWORDS

cardiac anxiety; cognitive behavioral therapy; health-related quality of life; internet delivered; noncardiac chest pain; psychological distress

Introduction

Background

Approximately 50% of patients seeking acute care because of chest pain have noncardiac chest pain (NCCP) [1-5], which is equivalent to approximately 80,000 people annually in Sweden [6]. Despite recurrent chest pain, many patients are discharged without a clear explanation regarding the underlying cause of their chest pain [7,8], and a large proportion of them continue to experience chest pain [8-10]. Wertli et al [11], in a retrospective study of 1341 emergency department admissions among patients with NCCP, concluded that there is an uncertainty among physicians regarding how to approach these patients after ruling out the cardiac causes. In fact, many of these patients are still not convinced by their negative cardiac diagnosis several years after the primary assessment, and they worry about heart disease, avoid activities that they think might be harmful to their heart, and repeatedly seek medical help [12,13]. Physical activity has been shown to be associated with a lower risk of experiencing NCCP. Therefore, avoidance of such activities has negative effects on these patients [14]. This situation seems to cause functional impairment, psychological distress, and impaired quality of life in these patients [4,15-18]. Studies have also suggested that patients with NCCP use outpatient health care to the same extent as patients with myocardial infarction [4,17,19] and incur high costs for the health care system and society [1,6,20-22].

Cardiac anxiety has been found to be highly prevalent in these patients and is strongly associated with health care use [9,18]. An explanation is that cardiac anxiety may worsen the chest pain and create a *vicious circle*, leading to the maintenance of both anxiety and chest pain. Psychological interventions targeting anxiety can be a feasible way of breaking the vicious circle and improving patient outcomes by supporting patients in learning how to perceive and handle their chest pain. This is also warranted by the World Health Organization, which emphasizes the need for the implementation of interventions targeting psychological distress in patients with somatic diseases [23].

There is strong support for cognitive behavioral therapy (CBT) in the treatment of mild and moderately severe states of anxiety and depressive disorders [24-27]. CBT is a structured and collaborative process aiming to help individuals evaluate the accuracy and usefulness of their thoughts [28]. A Cochrane review found CBT useful and moderately successful, despite the multifaceted etiology in NCCP, and stressed the need for further randomized controlled trials of psychological interventions for NCCP with long follow-up periods [29]. However, a problem is that there are some barriers to the use of face-to-face CBT [30]. In addition to being time consuming, there is a lack of trained health care professionals who can deliver CBT, leading to a treatment demand gap; thus, the number of patients in need of CBT is higher than the number being offered CBT [29,31]. A possible solution is to use the internet to deliver CBT (internet-delivered CBT [iCBT]), as it does not differ from face-to-face CBT regarding the effects among young and middle-aged people [32-35], requires less therapist involvement, can be given to more patients, is cheaper, and is not time or room dependent [30].

However, few studies have evaluated the effects and experiences of iCBT in patients with both somatic disease and psychological distress [36,37]. However, in a recent study, iCBT provided by nurses, with a brief introduction in iCBT, to patients with cardiac disease and depressive symptoms was shown to reduce depressive symptoms and improve the quality of life [38]. This implies that health care professionals, such as nurses within somatic health care, could use iCBT as a tool for treating patients, which could increase the access to CBT for psychological distress among patients with somatic diseases, such as NCCP. Furthermore, no studies have been undertaken concerning patients with NCCP, apart from our previous pilot study that tested the feasibility of iCBT in patients with NCCP and cardiac anxiety. Our piloted iCBT program was perceived as feasible by the patients [39]. Thus, there is a knowledge gap regarding the effects of iCBT programs guided by health care professionals, such as nurses, on cardiac anxiety and other patient-reported outcomes in patients with NCCP.

Objective

The aim of this study is to evaluate the effects of a nurse-led iCBT program on cardiac anxiety and other patient-reported outcomes in patients with NCCP.

Methods

Design and Ethics

This was a randomized controlled trial and was registered at ClinicalTrials.gov (NCT03336112). The study was approved by the regional ethical review board in Linköping, Sweden (code 2017/343-31).

Study Participants

Patients were eligible for participation if they were aged ≥18 years, had at least two health care consultations because of NCCP (International Classification of Diseases-10 codes R07.2, R07.3, R07.4, and Z03.4) during the past 6 months, and screened positive for cardiac anxiety (score ≥24 on the Cardiac Anxiety Questionnaire [CAQ]). Patients were excluded if they had no access to a computer or tablet with internet connection; were not able to perform physical activity because of physical constraints; did not speak or understand Swedish; or experienced severe depression, as measured by the Patient Health Questionnaire-9 (PHQ-9), cognitive impairment, or newly diagnosed cancer requiring treatment (according to medical records).

Procedures

Study participants were recruited after discharge from the emergency units at 3 regional hospitals and 1 university hospital in southeast Sweden. Recruitment of patients to the study was conducted between January 2018 and August 2020. At the start of the study, we only recruited patients from 2 hospitals; however, to allow for faster recruitment, we expanded to 2 more hospitals from April 2019. Eligible participants were identified by registers and sent a packet, which included study information, an informed consent form, and a prestamped envelope by post. After some days, patients were contacted by phone by the study team and informed verbally about the study. Patients interested in participation and who sent back a signed written informed consent form were screened for cardiac anxiety using an encrypted web-based survey tool provided by Linköping University and requiring both username and password to log in. Patients fulfilling the criteria were randomized in a 1:1 ratio into either iCBT or psychoeducation using a randomization table provided by a statistician. Masking of patients was not possible as the intervention was a CBT program. Participating in the study (in both the iCBT and psychoeducation groups) did not oppose seeking or receiving other inpatient or outpatient treatments.

The intervention was delivered through a website specifically developed for the study by the first author (GM). The website was accessible only to those who received the website URL and log-in details. In addition to username and password, 2-factor authentication was applied using an SMS text messaging one-time password to ensure security.

iCBT Group

The iCBT group received a 5-session/week–long nurse-led iCBT program that was developed based on our previous pilot study [39]. This pilot study contained 4 sessions and was extended by 1 week to give patients more time to work on the content for better effects. The program comprised psychoeducation, mindfulness, and exposure to physical activity, with weekly homework assignments. The psychoeducation part was aimed at teaching patients about chest pain and its impact on daily life and how avoidance and safety behaviors can maintain or exacerbate chest pain. A chest pain diary and exposure plan were set up by the patients to help them learn about their chest pain and not be afraid of exposing themselves to physical activity that many patients normally avoid. The mindfulness part, comprising both information and different exercises to perform daily, was intended to raise awareness of what is going on in the body, emotions, and sensations and to be more in the present despite chest pain to learn how to handle the chest pain. The physical activity part comprised information and recommendations regarding physical activity based on national guidelines [40]. The goal of this part was to enable patients to learn that their hearts tolerate physical activity and to reduce cardiac anxiety and avoidance of physical activity.

The patients had various weekly assignments to accomplish and send in for feedback. They then received feedback and advice from the same nurse once per week at the same time to allow them to plan their time and engagement in the program. The participants were fully aware of this procedure. The entire treatment was conducted through the study website. Reminders and encouraging messages were sent to motivate the patients to complete the program. The iCBT program accounted for the main part of the treatment. The guidance and feedback part took approximately a mean of 8 (SD 4) minutes per patient and week.

Psychoeducation Group

Patients with NCCP can be seen as a forgotten group as they usually have no regular follow-ups regarding their chest pain. Thus, their care as usual is equivalent to no care at all. Therefore, comparing iCBT with care as usual is equivalent to comparing with nothing, and therefore, the effect of iCBT could be overestimated, as it would be difficult to evaluate if the effect only depended on iCBT or partly on the attention of being in a study [41]. It is also recommended that control patients in iCBT studies receive an active control [36]. Furthermore, it is, for ethical reasons, better to offer patients with different problems some kind of intervention than nothing at all when randomizing them to the control arm [41]. On the basis of this, our active control group received psychoeducation (ie, containing the same information that the iCBT group received as part of their psychoeducation but without any assignments or feedback). This information was divided into 5 sessions, and patients accessed 1 session per week through the same website as the iCBT group. The active control group received a message from the therapist every week when a new session was made available.

Data Collection

Data were collected at baseline before randomization; at the end of the intervention (ie, 5 weeks from baseline), and at 3, 6, and 12 months after the end of the intervention. In this study, data from the 6- and 12-month follow-ups were not included. Medical data were provided via medical records, and the rest were self-reported. All self-reported data were collected using the encrypted web-based survey tool provided by Linköping University. As all questions were mandatory to be able to submit the questionnaires, this resulted in no missing values among those who chose to answer the questionnaires. In total, participants received 2 reminders every 2 weeks if they did not complete the questionnaires.

The primary outcome was cardiac anxiety. The secondary outcomes were fear of bodily sensations, depressive symptoms, health-related quality of life (HRQoL), chest pain frequency, health care use, and health care costs. Data on health care use and health care costs will be published separately.

Data Measurement

Cardiac Anxiety

Cardiac anxiety was measured by the CAQ. This measurement comprises 18 items, with scores ranging from 0 to 72. Higher scores indicate greater cardiac anxiety [42]. In this study, we used a median score of 24 from one of our previous studies [18] as the cutoff score for cardiac anxiety. In addition to the total score obtained from the 18 items, the CAQ comprises three subscales: fear, avoidance, and heart-focused attention. However, these subscales are not presented in this study. The total scale has demonstrated adequate reliability and validity

[42]. Cronbach α coefficients of .79 to .89 were confirmed in this study at the different measurement points.

Fear of Bodily Sensations

The Body Sensations Questionnaire (BSQ) was used to assess the fear of bodily sensations, such as palpitations, dizziness, and sweating. The BSQ comprises 17 items, with scores between 17 and 85. Higher scores indicate greater fear of bodily sensations [43]. The BSQ has been proven to be reliable and valid [13,18,43]. In this study, the Cronbach α coefficients were .92 to .94.

Depressive Symptoms

The prevalence and severity of depressive symptoms were assessed using the PHQ-9. This measurement comprises 9 items, with a score range between 0 and 27. Scores between 5 and 9 indicate mild depressive symptoms, 10 and 14 indicate moderate depressive symptoms, 15 and 19 indicate moderately severe depressive symptoms, and 20 and 27 indicate severe depressive symptoms. In this study, we used a cutoff score of 10 for depressive symptoms. The PHQ-9 is a reliable instrument [44], and in this study, the Cronbach α coefficients were .87 to .89.

Chest Pain Frequency

Data on chest pain frequency were collected with the following self-developed question: *During the last month, how often have you experienced NCCP?*

HRQoL Measure

The EuroQol visual analog scale (EQ-VAS), which is a part of EuroQol, was used to measure HRQoL. This EQ-VAS scores range between 0 (worst imaginable health state) and 100 (best imaginable health state). The EQ-VAS provides a quantitative measure of the patient's perception of their overall health and is a frequently used instrument [45].

Statistical Analysis

The SPSS (version 25; IBM Corp) was used for data analysis. The level of $P<.05$ was set for significance. According to our power calculation that was based on results from our previous pilot study [39], 53 participants needed to be included in each group to reach a 20% improvement (approximately an effect size of 0.5) in cardiac anxiety (95% CI and 80% power). This was used as there is no established score for a clinically significant improvement in this measurement. This sample size is comparable with similar CBT studies [32]. Intention-to-treat analysis was applied, and patients were followed up even if they were inactive or dropped out from the study. This method was chosen to preserve the integrity of the randomization and minimize the risks of bias related to differences in groups following attrition or nonadherence [46].

Frequencies, percentages, mean values, and SDs were used to describe and compare the study variables. Chi-square test or 2-tailed Student *t* test was used depending on the level of the data to compare differences between iCBT and psychoeducation groups regarding demographic variables.

To compare the iCBT and psychoeducation groups regarding changes in cardiac anxiety, bodily sensations, depressive symptoms, HRQoL, and NCCP frequency over the 3 different measurement points (ie, baseline and 5-week and 3-month follow-ups), a mixed model analysis was performed. As we had little missing data (in total, 8% at 5-week follow-up and 16% at 3-month follow-up), we chose to base our analysis on the original data; however, we also ran a mixed model analysis based on multiple imputation as a sensitivity analysis to ensure the accuracy of our results. Multiple imputation was based on the consideration of data being missing at random (the Little missing completely at random test; $P=.99$) [47]. According to this method, missing values depend on the observed data and are replaced by values based on the complete data set [48]. In this model, a total of 40 imputations were calculated based on the outcome and demographic variables that showed a significant correlation with the primary outcome at baseline. The results from the mixed model analysis based on multiple imputation did not differ from the analysis that was based on the original data; therefore, these results are not presented in this study.

Relationships between demographic variables, disease and psychological burden, treatment activity (such as number of log-ins and sessions performed), and change score in cardiac anxiety between baseline and 3-month follow-up in the groups were determined using the Pearson correlation coefficient. As there were no significant correlations, no further regression analysis was performed.

Cohen *d* was used to measure the effect sizes of the intervention. Effect sizes <0.20 were considered trivial, 0.20 to 0.49 were considered small, 0.50 to 0.79 were considered moderate, and ≥0.80 were considered large [49]. As there is no established score for a clinically significant improvement in CAQ, we calculated a reliable change index score according to the study by Christensen et al [50] using the baseline SD of 8.6 and baseline reliability score of 0.79 in our groups. On the basis of these values, a reliable change was deemed to be a change score of approximately 11 points per participant. The chi-square test was used to examine the difference between groups in the number of patients with a change score ≥11.

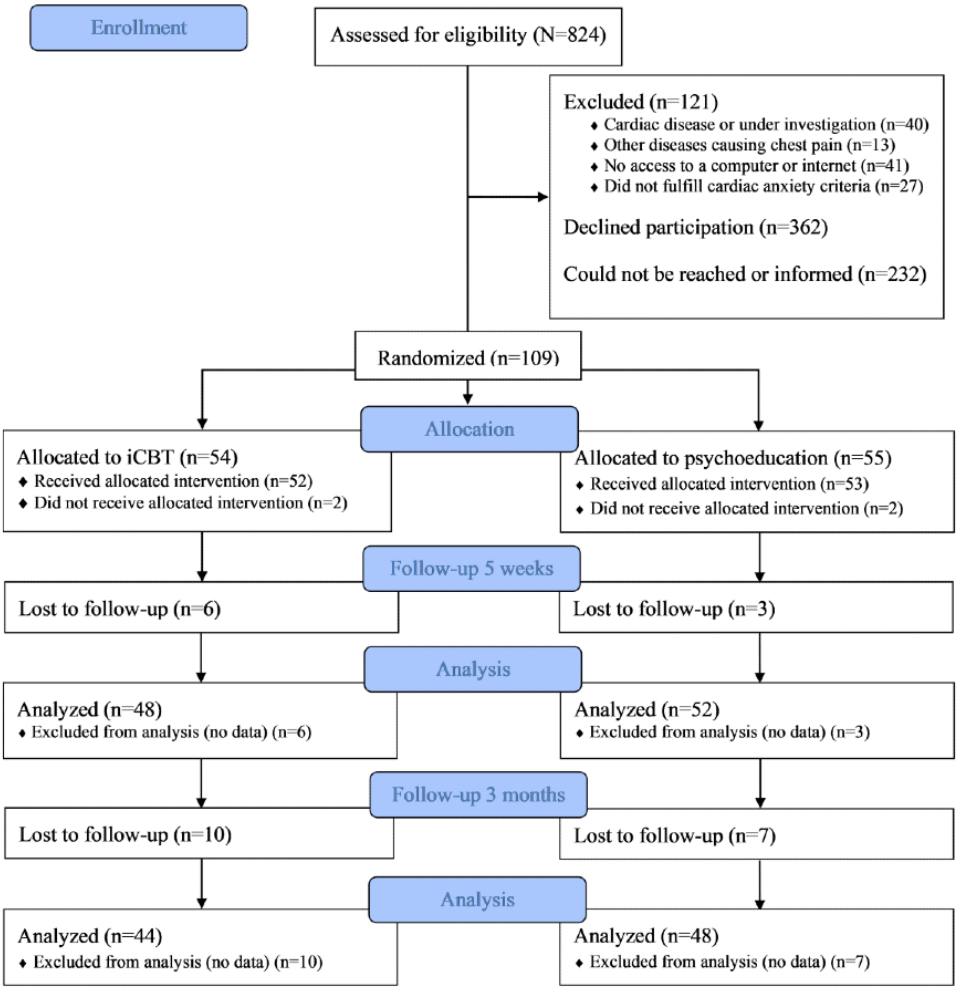
Within-group differences over the 3 different measurement points were analyzed using paired *t* tests.

Results

Study Participants

In total, 824 patients were assessed for eligibility and invited to participate in the study (Figure 1). After screening, 13.2% (109/824) of patients who fulfilled the inclusion and exclusion criteria and consented to participate were randomized into iCBT (54/109, 49.5%) or psychoeducation (55/109, 50.4%). Study participants included a higher proportion of women than nonparticipants (715/824, 86.8%; 67/109, 61.5% vs 354/715, 49.5%; $P=.02$).

Figure 1. The CONSORT (Consolidated Standards of Reporting Trials) flowchart. iCBT: internet-delivered cognitive behavioral therapy.



The demographic data for the iCBT and psychoeducation groups are presented in Table 1.

Patients in these groups had a mean age of 54 (17) years versus 57 (SD 16) years and were mainly women (32/54, 59% vs 35/55, 64%) and in a relationship (45/54, 83% vs 38/55, 69%). Approximately 33% (18/54) in the iCBT group and 27% (15/55) in the psychoeducation group had previous heart disease, 43% (23/54) in the iCBT group, and 34% (19/55) in the psychoeducation group reported previous problems with psychological distress (such as anxiety, panic, or depression), and more than half reported troubles with musculoskeletal pain. Furthermore, approximately half of them (iCBT 23/54, 43% vs psychoeducation 25/55, 45%) had been treated with psychotropic drugs. No significant differences were found between the groups for any of the demographic variables.

Among the participants, the attrition rates at the 5-week follow-up were 11% and 5% for the iCBT and psychoeducation groups, respectively. Patients in the iCBT group who were lost to the 5-week follow-up, that is, did not fill in the questionnaires, were more often on sick leave (P=.02). In addition, they had low or no engagement in the program; that is, they completed a mean of 1 session compared with 4.2 in those who continued their participation (P<.001). In the psychoeducation group, those lost to follow-up were more often single and living alone (P=.008). As for the iCBT group, those in the psychoeducation group who were lost to the 5-week follow-up had low or no engagement in the program, that is, completed a mean of 0.33 sessions compared with 4.0 in those who continued their participation (P<.001). The corresponding attrition rates of iCBT and psychoeducation groups for the 3-month follow-up were 19% and 13%, respectively, compared with baseline. The only differences between those who continued their participation and those lost to the 3-month follow-up was a lower number of completed sessions in the latter group (a mean of 4.1 vs 2.8 sessions, P=.02, in the iCBT group and a mean of 4.2 vs 1.6 sessions, P<.001, in the psychoeducation group).

Table 1. Demographic data of study patients at baseline (N=109).

Characteristics	iCBT[a] (n=54)	Psychoeducation (n=55)	*P* value
Age (year), mean (SD)	54.3 (16.5)	56.8 (15.5)	.42
Sex, n (%)			.64
Women	32 (59)	35 (64)	
Men	22 (41)	20 (36)	
Marital status, n (%)			.08
In a relationship	45 (83)	38 (69)	
Single	9 (17)	17 (31)	
Salary (US $), mean (SD)	3564 (5742)	4018 (5486)	.69
Economic situation, n (%)			.46
Very good	5 (9)	7 (13)	
Good	42 (78)	38 (69)	
Problematic	7 (13)	10 (18)	
Educational level, n (%)			.89
Compulsory school	7 (13)	8 (15)	
High school	23 (43)	25 (45)	
University	24 (44)	22 (40)	
Occupational status, n (%)			.30
Working	20 (37)	27 (49)	
Retired	20 (37)	19 (35)	
On sick leave	6 (11)	5 (9)	
Unemployed	2 (4)	3 (5)	
Student	6 (11)	1 (2)	
Smoking, n (%)			.97
Nonsmoker or previous smoker	49 (91)	48 (87)	
Smoker	5 (9)	7 (13)	
Alcohol consumption, n (%)			.58
Never or seldom	28 (52)	25 (45)	
≤9 glasses/week	24 (44)	30 (55)	
>9 glasses/week	2 (4)	0 (0)	
Exercise ≥30 minutes (days/week), mean (SD)	2.5 (2.1)	3.0 (2.2)	.24
Origin, n (%)			.36
Sweden	45 (85)	42 (76)	
Another Nordic country	2 (4)	3 (6)	
Another country within Europe	5 (9)	5 (9)	
South America	1 (2)	1 (2)	
Asia	0 (0)	4 (7)	
Charlson Comorbidity Index, mean (SD)	2.2 (2.4)	2.5 (2.1)	.57
Previous heart disease, n (%)	18 (33)	15 (27)	.49
Acid reflux, n (%)	9 (17)	9 (16)	.97
Muscle pain, n (%)	29 (54)	26 (47)	.50
Joint or skeletal pain, n (%)	32 (59)	31 (56)	.76
Psychological disorder, n (%)	23 (43)	19 (34)	.39

Characteristics	iCBT[a] (n=54)	Psychoeducation (n=55)	*P* value
Psychological treatment, n (%)	8 (15)	5 (9)	.36
Treatment with psychotropic drugs, n (%)	23 (43)	25 (45)	.76
Number of log-ins per week, mean (SD)	13.7 (12.8)	5.4 (2.5)	<.001
Sessions performed, n (%)			.06
0	2 (4)	2 (4)	
1	5 (9)	4 (7)	
2	7 (13)	4 (7)	
3	5 (9)	9 (16)	
4	1 (2)	10 (18)	
5	34 (63)	26 (47)	

[a]iCBT: internet-delivered cognitive behavioral therapy.

The Effects of iCBT Compared With Psychoeducation on Cardiac Anxiety

The mixed model analysis showed no significant differences between the iCBT and psychoeducation groups regarding cardiac anxiety in terms of the interaction effect of time and group over the 3-month follow-up (Table 2). However, as shown in Figure 2, the iCBT group continued to show improvements in cardiac anxiety at the 3-month follow-up, which was not the case for the psychoeducation group. The analysis showed a small effect size in favor of iCBT (Cohen *d*=0.31). Looking at the time effect, there were statistically significant differences in cardiac anxiety at both the 5-week and 3-month follow-ups compared with baseline (*P*<.001).

In the iCBT group, 36% (16/44) of patients had a positive reliable change score between 11 and 34 points, indicating improvement in cardiac anxiety, whereas 2% (1/44) of patients had a negative reliable change score, showing a deterioration in cardiac anxiety. The corresponding numbers in the psychoeducation group were 27% (13/48) of patients with a positive reliable change score and 2% (1/48) of patients with a negative reliable change score. However, the difference between the groups was nonsignificant (*P*=.21).

Within-group analysis showed that both iCBT and psychoeducation groups improved significantly with respect to cardiac anxiety from baseline to 5-week follow-up (*P*<.001). The iCBT group had an improved mean score from 35.8 to 29.0, whereas the psychoeducation group had an improved score from 36.3 to 30.8. In addition, at the 3-month follow-up, the mean score in the iCBT group was 26.6, which was significantly lower than both baseline (*P*<.001) and the 5-week follow-up (*P*=.04).

Table 2. Mixed model analysis of the effect of internet-delivered cognitive behavioral therapy (iCBT) compared with psychoeducation on cardiac anxiety and secondary outcomes, presented in estimated marginal means.

Variables	Time effect		Group effect			Interaction effect			Effect size (Cohen *d*)
	Value, mean	*P* value[a]	iCBT	Psychoeducation	*P* value	iCBT	Psychoeducation	*P* value[a]	
CAQ[b]			31.3	32.4	.96				
Baseline	36.4	N/A[c]				36.3	36.4	N/A	N/A
5 weeks	30.3	<.001				29.7	30.8	.46	N/A
3 months	28.9	<.001				27.9	29.9	.28	0.31
BSQ[d]			39.3	39.5	.62				
Baseline	42.7	N/A				43.3	42.1	N/A	N/A
5 weeks	38.1	<.001				37.1	39.1	.10	N/A
3 months	37.4	<.001				37.5	37.4	.55	0.15
PHQ-9[e]			7.2	6.6	.76				
Baseline	7.8	N/A				8.0	7.6	N/A	N/A
5 weeks	7.0	.54				7.6	6.4	.32	N/A
3 months	5.9	.005				6.0	5.8	.87	0.10
EQ-VAS[f]			60.4	63.4	.03				
Baseline	62.5	N/A				58.5	66.4	N/A	N/A
5 weeks	61.4	.49				60.4	62.3	.12	N/A
3 months	61.9	.32				62.5	61.5	.10	0.57
Chest pain frequency			13	7.9	.006				
Baseline	12.0	N/A				15.1	8.9	N/A	N/A
5 weeks	11.1	.60				14.1	8.1	.93	N/A
3 months	8.3	.005				9.7	6.8	.20	0.21

[a]In comparison with baseline.

[b]CAQ: Cardiac Anxiety Questionnaire.

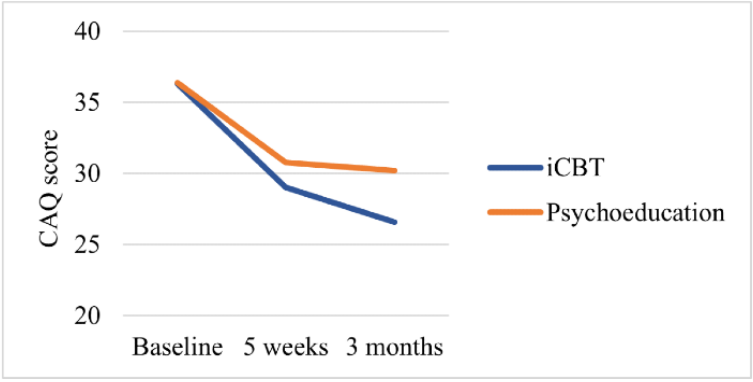
[c]N/A: not applicable.

[d]BSQ: Body Sensations Questionnaire.

[e]PHQ-9: Patient Health Questionnaire-9.

[f]EQ-VAS: EuroQol visual analog scale.

Figure 2. Changes in cardiac anxiety over time between internet-delivered cognitive behavioral therapy (iCBT) and psychoeducation groups. CAQ: Cardiac Anxiety Questionnaire.



The Effects of iCBT Compared With Psychoeducation on Secondary Outcomes

No significant interaction effect of time and group was found in any of the secondary outcomes (ie, fear of bodily sensations, depressive symptoms, HRQoL, and chest pain frequency) over the 3-month follow-up in the iCBT and psychoeducation groups (Table 2 and Figure 3). Table 3 displays the number of patients with depressive symptoms ≥10 in the iCBT and psychoeducation groups. In the iCBT group, only 11% (6/54) of patients reported depressive symptoms at the 3-month follow-up compared with 39% (21/54) of patients at baseline. For the psychoeducation group, of the 55 patients, the corresponding numbers were 9 (16%) patients at the 3-month follow-up and 13 (24%) at baseline. The effect sizes of the iCBT group on the secondary outcomes were trivial to moderate (Cohen *d* between 0.10 and 0.57) compared with those of the psychoeducation group. There were statistically significant differences regarding the time effect in bodily sensations at both 5-week and 3-month follow-ups ($P<.001$) and in depressive symptoms and chest pain frequency at the 3-month follow-up ($P=.005$) compared with baseline. Comparing the groups without considering the time effects, there were significant differences in HRQoL ($P=.03$) and chest pain frequency ($P=.006$) in favor of iCBT.

Figure 3. Changes in bodily sensations, depressive symptoms, EuroQol visual analog scale (EQ-VAS), and noncardiac chest pain (NCCP) frequency over time between internet-delivered cognitive behavioral therapy (iCBT) and psychoeducation groups. BSQ: Body Sensations Questionnaire; PHQ-9: Patient Health Questionnaire-9.

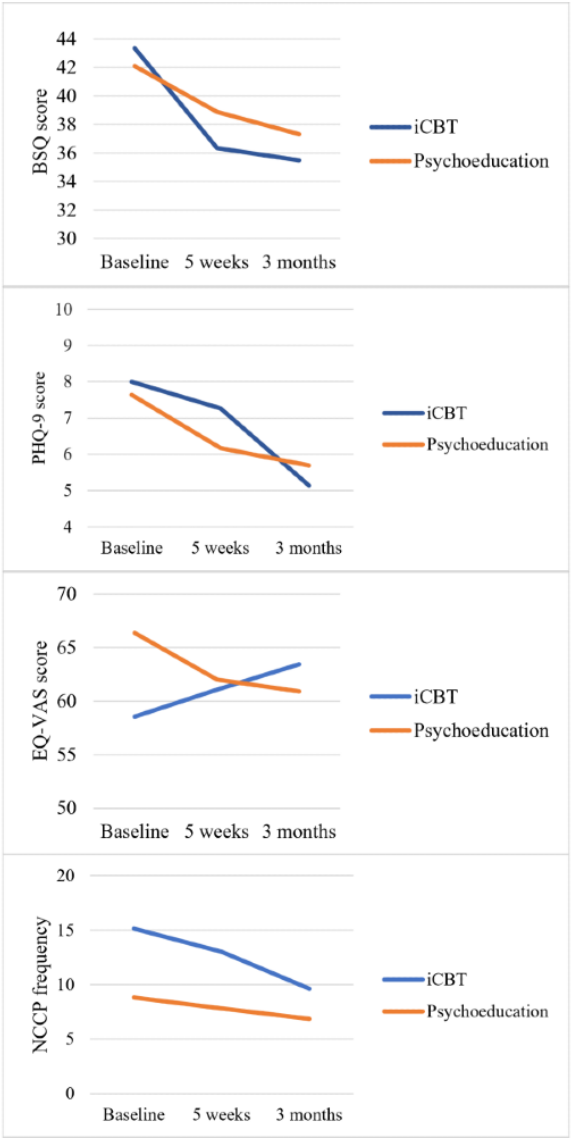


Table 3. Changes in the number of patients with depressive symptoms ≥10 between baseline and 3-month follow-up in the internet-delivered cognitive behavioral therapy (iCBT) and psychoeducation groups.

Measurement point	Patients with depressive symptoms, n (%)	
	iCBT (n=54)	Psychoeducation (n=55)
Baseline	21 (39)	13 (24)
5-week follow-up	15 (28)	9 (16)
3-month follow-up	6 (11)	9 (16)

Discussion

Principal Findings

To our knowledge, this is the first randomized controlled iCBT study targeting cardiac anxiety in patients with NCCP. In this study, we compared a 5-week–long nurse-led iCBT program comprising psychoeducation, mindfulness, and exposure to physical activity, with psychoeducation only to evaluate the effects on psychological distress, including cardiac anxiety, and other patient-reported outcomes in patients with NCCP. Our findings showed that iCBT was not superior to psychoeducation in decreasing cardiac anxiety in these patients. We found that one-third of the patients who received iCBT reported decreased anxiety levels compared with one-fourth of those who received psychoeducation, that is, a reliable change score of at least 11 points on the CAQ. Another important aspect is that only one of the patients in each group reported increased cardiac anxiety scores at the 3-month follow-up, indicating that the program was safe.

Although no significant interaction effect of time and group was found between the groups, iCBT seems to have better long-term effects on cardiac anxiety than psychoeducation, as within-group analysis showed that patients in the iCBT group had improved further at the 3-month follow-up (P=.04), which the psychoeducation group did not. In addition, patients who received iCBT tended to show long-term improvement in all reported variables (Figures 2 and 3), whereas those in the psychoeducation group did not change further. We also found a possible interaction effect of time and group regarding HRQoL in favor of the iCBT group with a moderate effect size. The patients improved over the measurement points (EQ-VAS increased from 58.5 at baseline to 65.5 at the 3-month follow-up), whereas the psychoeducation group deteriorated (EQ-VAS decreased from 66.4 at baseline to 61.5 at the 3-month follow-up).

These findings point to the fact that results related to changes in psychological distress and HRQoL may take more time to appear and that our intervention might be too short to detect such changes. A systematic review and meta-analysis by Reavell et al [51] showed a significant reduction in anxiety in long CBT interventions but not in short- to medium-term interventions compared with controls. These interventions were performed in patients with cardiovascular disease, and patients might need a longer time to adjust to their cardiac diagnosis and handle their anxiety compared with patients with NCCP. Previous studies in patients with NCCP using short interventions have reported inconsistent results [13,25,52,53]. Jonsbu et al [13] used a 3-session CBT intervention and reported significant effects in favor of CBT on fear of bodily sensations, avoidance of physical activity, depression, and some domains of HRQoL, and these effects lasted up to 12 months. However, cardiac anxiety was not studied. In a study by van Beek et al [25], positive effects on anxiety were found 24 weeks following a 6-session CBT intervention compared with usual treatment in patients with NCCP and panic disorder. Mulder et al [52] reported significantly lower health anxiety scores at 3 months but not at 12 months in patients who received 3 to 4 sessions of CBT compared with care as usual. The authors suggested booster sessions or longer interventions for more sustained improvement. Tyrer et al [53] failed to present any significant differences between iCBT and standard care at both 6- and 12-month follow-ups. Their CBT intervention comprised 4 to 10 sessions (mean 5.7) during a mean period of 14.3 weeks. They concluded that these results were mainly because of the study being underpowered. In an 8-year follow-up of CBT on health anxiety [54], the authors reported sustained improvement in health anxiety symptoms, especially among patients with cardiac diseases, and this improvement was greatest in nurse-led CBT. However, different length and follow-up periods for the interventions were used in these studies, which also used face-to-face CBT, compared with passive controls, which makes these studies difficult to compare with our study. This suggests that CBT is valuable to medical patients, including NCCP, with anxiety; however, the key is to find the right duration and type of CBT modality.

Furthermore, we believe that we could not find significant differences as iCBT was compared with an active control. In addition to the possible placebo effect of being in a trial and having their symptoms recognized, receiving psychoeducation, which our control group did, seems to have had effects on both primary and secondary outcomes, which could explain the minor and nonsignificant differences between the iCBT and psychoeducation groups [41]. Psychoeducation makes it easier for patients to understand what therapeutic tools to use and facilitates the motivation for behavior change and finding different ways of thinking in relation to their condition [55]. This can be compared with other studies comparing iCBT with usual care or waiting list instead of an active control and reporting significant improvement in anxiety and depression in favor of iCBT [25,30]. Studies with passive controls, such as waiting list or standard care, often report better outcomes in favor of the intervention compared with those with active controls [56]. An important aspect to consider is how long the effects of psychoeducation last and whether more psychological tools, such as cognitive restructuring and behavioral change strategies, are needed to maintain long-term results.

In the iCBT group, a larger proportion (21/54, 39% vs 13/55, 24%) of patients had at least moderate levels of depressive symptoms at baseline, although the mean values in the group did not differ significantly from those in the psychoeducation group. Although the study was not designed to target depressive symptoms, the iCBT program had significant effects on depressive symptoms in most of these patients, as 71% (15/21) reported lower scores on the PHQ-9 (ie, <10) than was reported in patients who received psychoeducation, in which only 31% (4/13) reduced depressive symptom levels. On the basis of our previous findings revealing that depressive symptoms have strong effects on cardiac anxiety in patients with NCCP, we believe that the effects of iCBT targeting cardiac anxiety might have been limited by the higher proportion of patients with depressive symptoms in this group [57]. Another affecting aspect could be that the length of our program was too short to lead to improvement, as comorbid depressive symptoms need longer treatments than if patients only have anxiety [58].

Strengths and Limitations

Of the 824 patients assessed, only 109 (13.2%) took part in the study. This could indicate that the population in this study may not be representative of patients with NCCP in general. Patients were included based on high levels of anxiety, as the program was designed to target anxiety. As the CAQ lacks a cutoff score for anxiety, we used the median value in our previous study. This might have been too high, as many patients who perceived themselves to have cardiac anxiety did not come up to this cutoff and were, therefore, excluded from the study. As we had little missing values, we chose to report the original data; however, we also provided a mixed model analysis with multiple imputed data to ensure the accuracy of our results. Regardless of the type of statistical analysis, our result proved to remain. Some patients could not join the study because of a lack of computer or internet access or perceive themselves as having difficulties in handling such interventions. However, we are aware that internet interventions cannot fit all patients, partly because of technical aspects but mainly because of preferences for face-to-face therapies [30]. We had difficulties recruiting patients to the study and had to expand our recruitment sites during the study. This could be as many patients with NCCP are not motivated to receive psychological treatment as they perceive their pain to be physical. As stated previously, the iCBT program accounted for the main part of the treatment effect, and the guidance and feedback part mainly aimed at confirming patients' work and progress and encouraging them to reflect on different issues and take on new challenges. The guidance and feedback were mainly provided by the first author (GM) and, in some cases, by the second author (MEL), who was mentored by the first author. This indicates that feedback can be provided by different health care professionals with a short introduction to iCBT.

Conclusions

iCBT was not superior to psychoeducation in decreasing cardiac anxiety in patients with NCCP. However, iCBT tends to have better long-term effects on psychological distress, including cardiac anxiety, HRQoL, and NCCP frequency than psychoeducation. Better conclusions may be drawn based on further follow-ups.

Acknowledgments

The authors would like to acknowledge the Kamprad Family Foundation that funded the study (reference number ISV-2018-00033). The funder had no role in any part of the study.

Authors' Contributions

GM, PK, and PJ contributed to the conception and design of the study. GM, MEL, and PK collected data. GM, MEL, and PJ performed statistical analyses and interpretation of the data and contributed to the drafting of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
CONSORT-eHEALTH (V1.6.1)
[PDF File (Adobe PDF File), 722 KB - jmir_v24i1e31674_app1.pdf]

References

1. Mol KA, Smoczynska A, Rahel BM, Meeder JG, Janssen L, Doevendans PA, et al. Non-cardiac chest pain: prognosis and secondary healthcare utilisation. Open Heart 2018 Oct 15;5(2):e000859 [FREE Full text] [doi: 10.1136/openhrt-2018-000859] [Medline: 30364505]
2. McDevitt-Petrovic O, Kirby K, Shevlin M. The prevalence of non-cardiac chest pain (NCCP) using emergency department (ED) data: a Northern Ireland based study. BMC Health Serv Res 2017 Aug 09;17(1):549 [FREE Full text] [doi: 10.1186/s12913-017-2493-8] [Medline: 28793910]
3. Campbell KA, Madva EN, Villegas AC, Beale EE, Beach SR, Wasfy JH, et al. Non-cardiac chest pain: a review for the consultation-liaison psychiatrist. Psychosomatics 2017 May;58(3):252-265 [FREE Full text] [doi: 10.1016/j.psym.2016.12.003] [Medline: 28196622]

4. Webster R, Norman P, Goodacre S, Thompson A. The prevalence and correlates of psychological outcomes in patients with acute non-cardiac chest pain: a systematic review. Emerg Med J 2012 Apr 27;29(4):267-273. [doi: 10.1136/emermed-2011-200526] [Medline: 22034535]
5. Bjørnsen LP, Naess-Pleym LE, Dale J, Grenne B, Wiseth R. Description of chest pain patients in a Norwegian emergency department. Scand Cardiovasc J 2019 Feb 26;53(1):28-34. [doi: 10.1080/14017431.2019.1583362] [Medline: 30760035]
6. Mourad G, Alwin J, Strömberg A, Jaarsma T. Societal costs of non-cardiac chest pain compared with ischemic heart disease--a longitudinal study. BMC Health Serv Res 2013 Oct 09;13(1):403 [FREE Full text] [doi: 10.1186/1472-6963-13-403] [Medline: 24107009]
7. Reid S, Wessely S, Crayford T, Hotopf M. Medically unexplained symptoms in frequent attenders of secondary health care: retrospective cohort study. Br Med J 2001 Mar 31;322(7289):767 [FREE Full text] [doi: 10.1136/bmj.322.7289.767] [Medline: 11282861]
8. Spalding L, Reay E, Kelly C. Cause and outcome of atypical chest pain in patients admitted to hospital. J R Soc Med 2003 Mar 01;96(3):122-125 [FREE Full text] [doi: 10.1258/jrsm.96.3.122] [Medline: 12612112]
9. Tremblay M, Denis I, Turcotte S, Fleet RP, Archambault P, Dionne CE, et al. Heart-focused anxiety and health care seeking in patients with non-cardiac chest pain: a prospective study. Gen Hosp Psychiatry 2018 Jan;50:83-89. [doi: 10.1016/j.genhosppsych.2017.10.007] [Medline: 29120733]
10. TEW R, Guthrie EA, Creed FH, Cotter L, Kisely S, Tomenson B. A long-term follow-up study of patients with ischaemic heart disease versus patients with nonspecific chest pain. J Psychosom Res 1995 Nov;39(8):977-985. [doi: 10.1016/0022-3999(95)00065-8]
11. Wertli MM, Dangma TD, Müller SE, Gort LM, Klauser BS, Melzer L, et al. Non-cardiac chest pain patients in the emergency department: do physicians have a plan how to diagnose and treat them? A retrospective study. PLoS One 2019 Feb 1;14(2):e0211615 [FREE Full text] [doi: 10.1371/journal.pone.0211615] [Medline: 30707725]
12. Bass C, Mayou R. Chest pain. Br Med J 2002 Sep 14;325(7364):588-591 [FREE Full text] [doi: 10.1136/bmj.325.7364.588] [Medline: 12228139]
13. Jonsbu E, Dammen T, Morken G, Moum T, Martinsen EW. Short-term cognitive behavioral therapy for non-cardiac chest pain and benign palpitations: a randomized controlled trial. J Psychosom Res 2011 Feb;70(2):117-123. [doi: 10.1016/j.jpsychores.2010.09.013] [Medline: 21262413]
14. Castonguay J, Turcotte S, Fleet RP, Archambault PM, Dionne CE, Denis I, et al. Physical activity and disability in patients with noncardiac chest pain: a longitudinal cohort study. Biopsychosoc Med 2020 Jun 30;14(1):12 [FREE Full text] [doi: 10.1186/s13030-020-00185-9] [Medline: 32612673]
15. Goodacre S, Mason S, Arnold J, Angelini K. Psychologic morbidity and health-related quality of life of patients assessed in a chest pain observation unit. Ann Emerg Med 2001 Oct;38(4):369-376. [doi: 10.1067/mem.2001.118010] [Medline: 11574792]
16. Eslick G, Talley NJ. Non-cardiac chest pain: predictors of health care seeking, the types of health care professional consulted, work absenteeism and interruption of daily activities. Aliment Pharmacol Ther 2004 Oct 15;20(8):909-915 [FREE Full text] [doi: 10.1111/j.1365-2036.2004.02175.x] [Medline: 15479363]
17. Mourad G, Jaarsma T, Hallert C, Strömberg A. Depressive symptoms and healthcare utilization in patients with noncardiac chest pain compared to patients with ischemic heart disease. Heart Lung 2012 Sep;41(5):446-455. [doi: 10.1016/j.hrtlng.2012.04.002] [Medline: 22652167]
18. Mourad G, Strömberg A, Johansson P, Jaarsma T. Depressive symptoms, cardiac anxiety, and fear of body sensations in patients with non-cardiac chest pain, and their relation to healthcare-seeking behavior: a cross-sectional study. Patient 2016 Feb 4;9(1):69-77. [doi: 10.1007/s40271-015-0125-0] [Medline: 25840677]
19. Eslick GD. Health care seeking behaviors, psychological factors, and quality of life of noncardiac chest pain. Dis Mon 2008 Sep;54(9):604-612. [doi: 10.1016/j.disamonth.2008.06.004] [Medline: 18725002]
20. Eslick G, Coulshed DS, Talley NJ. Review article: the burden of illness of non-cardiac chest pain. Aliment Pharmacol Ther 2002 Jul;16(7):1217-1223 [FREE Full text] [doi: 10.1046/j.1365-2036.2002.01296.x] [Medline: 12144570]
21. Forberg JL, Henriksen LS, Edenbrandt L, Ekelund U. Direct hospital costs of chest pain patients attending the emergency department: a retrospective study. BMC Emerg Med 2006 May 04;6(1). [doi: 10.1186/1471-227x-6-6]
22. Groarke J, O'Brien J, Go G, Susanto M, Owens P, Maree AO. Cost burden of non-specific chest pain admissions. Ir J Med Sci 2013 Mar 3;182(1):57-61. [doi: 10.1007/s11845-012-0826-5] [Medline: 22552895]
23. Moussavi S, Chatterji S, Verdes E, Tandon A, Patel V, Ustun B. Depression, chronic diseases, and decrements in health: results from the World Health Surveys. Lancet 2007 Sep;370(9590):851-858. [doi: 10.1016/s0140-6736(07)61415-9]
24. The Swedish National Board of Health and Welfare. Vård vid depression och ångestsyndrom. Stöd för styrning och ledning. 2019. URL: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/nationella-riktlinjer/2019-5-13.pdf [accessed 2022-01-13]
25. van Beek M, Voshaar R, Beek A, van Zijderveld G, Visser S, Speckens A, et al. A brief cognitive-behavioral intervention for treating depression and panic disorder in patients with noncardiac chest pain: a 24-week randomized controlled trial. Depress Anxiety 2013 Jul 26;30(7):670-678. [doi: 10.1002/da.22106] [Medline: 23625592]

26. Høifødt RS, Strøm C, Kolstrup N, Eisemann M, Waterloo K. Effectiveness of cognitive behavioural therapy in primary health care: a review. Fam Pract 2011 Oct 09;28(5):489-504. [doi: 10.1093/fampra/cmr017] [Medline: 21555339]
27. Hofmann SG, Asnaani A, Vonk IJ, Sawyer AT, Fang A. The efficacy of cognitive behavioral therapy: a review of meta-analyses. Cognit Ther Res 2012 Oct 01;36(5):427-440 [FREE Full text] [doi: 10.1007/s10608-012-9476-1] [Medline: 23459093]
28. Beck AT, Dozois DJ. Cognitive therapy: current status and future directions. Annu Rev Med 2011 Feb 18;62(1):397-409. [doi: 10.1146/annurev-med-052209-100032] [Medline: 20690827]
29. Kisely S, Campbell LA, Skerritt P. Psychological interventions for symptomatic management of non-specific chest pain in patients with normal coronary anatomy. Cochrane Database Syst Rev 2005 Jan 25;2015(1):CD004101. [doi: 10.1002/14651858.CD004101.pub2] [Medline: 15674930]
30. Health Quality Ontario. Internet-delivered cognitive behavioural therapy for major depression and anxiety disorders: a health technology assessment. Ont Health Technol Assess Ser 2019;19(6):1-199 [FREE Full text] [Medline: 30873251]
31. The Swedish National Board of Health and Welfare. Långvarig psykisk ohälsa en utmaning för vården. Socialstyrelsen. 2019. URL: https://www.socialstyrelsen.se/om-socialstyrelsen/pressrum/press/langvarig-psykisk-ohalsa-en-utmaning-for-varden/ [accessed 2022-01-13]
32. Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N. Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One 2010 Oct 13;5(10):e13196 [FREE Full text] [doi: 10.1371/journal.pone.0013196] [Medline: 20967242]
33. Johansson R, Andersson G. Internet-based psychological treatments for depression. Expert Rev Neurother 2012 Jul 09;12(7):861-870. [doi: 10.1586/ern.12.63] [Medline: 22853793]
34. Carlbring P, Andersson G, Cuijpers P, Riper H, Hedman-Lagerlöf E. Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: an updated systematic review and meta-analysis. Cogn Behav Ther 2018 Jan 07;47(1):1-18. [doi: 10.1080/16506073.2017.1401115] [Medline: 29215315]
35. Olthuis J, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2015 Mar 05;3(3):CD011565. [doi: 10.1002/14651858.CD011565] [Medline: 25742186]
36. McCombie A, Gearry R, Andrews J, Mikocka-Walus A, Mulder R. Computerised cognitive behavioural therapy for psychological distress in patients with physical illnesses: a systematic review. J Clin Psychol Med Settings 2015 Mar 11;22(1):20-44. [doi: 10.1007/s10880-015-9420-0] [Medline: 25666485]
37. Mehta S, Peynenburg VA, Hadjistavropoulos HD. Internet-delivered cognitive behaviour therapy for chronic health conditions: a systematic review and meta-analysis. J Behav Med 2019 Apr 01;42(2):169-187. [doi: 10.1007/s10865-018-9984-x] [Medline: 30387008]
38. Johansson P, Westas M, Andersson G, Alehagen U, Broström A, Jaarsma T, et al. An internet-based cognitive behavioral therapy program adapted to patients with cardiovascular disease and depression: randomized controlled trial. JMIR Ment Health 2019 Oct 03;6(10):e14648 [FREE Full text] [doi: 10.2196/14648] [Medline: 31584000]
39. Mourad G, Strömberg A, Jonsbu E, Gustafsson M, Johansson P, Jaarsma T. Guided Internet-delivered cognitive behavioural therapy in patients with non-cardiac chest pain - a pilot randomized controlled study. Trials 2016 Jul 26;17(1):352 [FREE Full text] [doi: 10.1186/s13063-016-1491-1] [Medline: 27456689]
40. Physical activity in the prevention and treatment of diseases. Swedish National Institute of Public Health. 2011. URL: https://www.folkhalsomyndigheten.se/publicerat-material/publikationsarkiv/p/physical-activity-in-the-prevention-and-treatment-of-disease/ [accessed 2022-01-13]
41. Aycock DM, Hayat MJ, Helvig A, Dunbar SB, Clark PC. Essential considerations in developing attention control groups in behavioral research. Res Nurs Health 2018 Jun 14;41(3):320-328. [doi: 10.1002/nur.21870] [Medline: 29906317]
42. Eifert GH, Thompson RN, Zvolensky MJ, Edwards K, Frazer NL, Haddad JW, et al. The Cardiac Anxiety Questionnaire: development and preliminary validity. Behav Res Ther 2000 Oct;38(10):1039-1053. [doi: 10.1016/s0005-7967(99)00132-1]
43. Chambless DL, Caputo GC, Bright P, Gallagher R. Assessment of fear of fear in agoraphobics: The Body Sensations Questionnaire and the Agoraphobic Cognitions Questionnaire. J Consult Clin Psychol 1984 Dec;52(6):1090-1097. [doi: 10.1037/0022-006x.52.6.1090]
44. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med 2001 Sep;16(9):606-613 [FREE Full text] [doi: 10.1046/j.1525-1497.2001.016009606.x] [Medline: 11556941]
45. EQ-5D. EuroQolGroup. 2017. URL: https://euroqol.org/ [accessed 2022-01-13]
46. McCoy E. Understanding the intention-to-treat principle in randomized controlled trials. West J Emerg Med 2017 Oct 18;18(6):1075-1078 [FREE Full text] [doi: 10.5811/westjem.2017.8.35985] [Medline: 29085540]
47. Jakobsen JC, Gluud C, Wetterslev J, Winkel P. When and how should multiple imputation be used for handling missing data in randomised clinical trials - a practical guide with flowcharts. BMC Med Res Methodol 2017 Dec 06;17(1):162 [FREE Full text] [doi: 10.1186/s12874-017-0442-1] [Medline: 29207961]
48. Little RA, Rubin DB. Statistical Analysis with Missing Data, Second Edition. Hoboken, New Jersey, United States: John Wiley & Sons; 2002.
49. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd Ed. New York: Routledge; 1988.

50. Christensen L, Mendoza JL. A method of assessing change in a single subject: an alteration of the RC index. Behav Ther 1986 Jun;17(3):305-308. [doi: 10.1016/s0005-7894(86)80060-0]
51. Reavell J, Hopkinson M, Clarkesmith D, Lane DA. Effectiveness of cognitive behavioral therapy for depression and anxiety in patients with cardiovascular disease: a systematic review and meta-analysis. Psychosom Med 2018;80(8):742-753. [doi: 10.1097/psy.0000000000000626]
52. Mulder R, Zarifeh J, Boden J, Lacey C, Tyrer P, Tyrer H, et al. An RCT of brief cognitive therapy versus treatment as usual in patients with non-cardiac chest pain. Int J Cardiol 2019 Aug 15;289:6-11. [doi: 10.1016/j.ijcard.2019.01.067] [Medline: 30770262]
53. Tyrer P, Tyrer H, Morriss R, Crawford M, Cooper S, Yang M, et al. Clinical and cost-effectiveness of adapted cognitive behaviour therapy for non-cardiac chest pain: a multicentre, randomised controlled trial. Open Heart 2017 May 16;4(1):e000582 [FREE Full text] [doi: 10.1136/openhrt-2016-000582] [Medline: 28674627]
54. Tyrer P, Wang D, Crawford M, Dupont S, Cooper S, Nourmand S, et al. Sustained benefit of cognitive behaviour therapy for health anxiety in medical patients (CHAMP) over 8 years: a randomised-controlled trial. Psychol Med 2020 Mar 16;51(10):1714-1722. [doi: 10.1017/s003329172000046x]
55. Borza L. Cognitive-behavioral therapy for generalized anxiety. Dialogues Clin Neurosci 2017 Jul;19(2):203-208. [doi: 10.31887/dcns.2017.19.2/lborza]
56. LaFave SE, Granbom M, Cudjoe TK, Gottsch A, Shorb G, Szanton SL. Attention control group activities and perceived benefit in a trial of a behavioral intervention for older adults. Res Nurs Health 2019 Dec 24;42(6):476-482 [FREE Full text] [doi: 10.1002/nur.21992] [Medline: 31647125]
57. Mourad G, Jaarsma T, Strömberg A, Svensson E, Johansson P. The associations between psychological distress and healthcare use in patients with non-cardiac chest pain: does a history of cardiac disease matter? BMC Psychiatry 2018 Jun 05;18(1):172 [FREE Full text] [doi: 10.1186/s12888-018-1689-8] [Medline: 29866125]
58. Liu J, Gill NS, Teodorczuk A, Li Z, Sun J. The efficacy of cognitive behavioural therapy in somatoform disorders and medically unexplained physical symptoms: a meta-analysis of randomized controlled trials. J Affect Disord 2019 Feb 15;245:98-112. [doi: 10.1016/j.jad.2018.10.114] [Medline: 30368076]

Abbreviations

BSQ: Body Sensations Questionnaire
CAQ: Cardiac Anxiety Questionnaire
CBT: cognitive behavioral therapy
EQ-VAS: EuroQol visual analog scale
HRQoL: health-related quality of life
iCBT: internet-delivered cognitive behavioral therapy
NCCP: noncardiac chest pain
PHQ-9: Patient Health Questionnaire-9

Edited by A Mavragani; submitted 30.06.21; peer-reviewed by P Tyrer, K Matsumoto; comments to author 04.11.21; revised version received 09.11.21; accepted 08.12.21; published 28.01.22.

Please cite as:
Mourad G, Eriksson-Liebon M, Karlström P, Johansson P
The Effect of Internet-Delivered Cognitive Behavioral Therapy Versus Psychoeducation Only on Psychological Distress in Patients With Noncardiac Chest Pain: Randomized Controlled Trial
J Med Internet Res 2022;24(1):e31674
URL: https://www.jmir.org/2022/1/e31674
doi:10.2196/31674
PMID:35089153

Original Paper

Factors Associated With the Experience of Cognitive Training Apps for the Prevention of Dementia: Cross-sectional Study Using an Extended Health Belief Model

Jaegyeong Lee[1], RN, MSN, PhD; Jung Min Lim[2], MD

[1]College of Nursing, Seoul National University, Seoul, Republic of Korea

[2]Department of Pharmacology, Yonsei University College of Medicine, Seoul, Republic of Korea

Corresponding Author:
Jung Min Lim, MD
Department of Pharmacology
Yonsei University College of Medicine
Avison Biomedical Research Center, Room 318
50-1 Yonsei-ro, Seodaemun-gu
Seoul, 03722
Republic of Korea
Phone: 82 2 2228 0925
Email: dreamist88@gmail.com

Abstract

Background: The prevalence and economic burden of dementia are increasing dramatically. Using information communication technology to improve cognitive functions is proven to be effective and holds the potential to serve as a new and efficient method for the prevention of dementia.

Objective: The aim of this study was to identify factors associated with the experience of mobile apps for cognitive training in middle-aged adults. We evaluated the relationships between the experience of cognitive training apps and structural variables using an extended health belief model.

Methods: An online survey was conducted on South Korean participants aged 40 to 64 years (N=320). General characteristics and dementia knowledge were measured along with the health belief model constructs. Statistical analysis and logistic regression analysis were performed.

Results: Higher dementia knowledge (odds ratio [OR] 1.164, *P*=.02), higher perceived benefit (OR 1.373, *P*<.001), female gender (OR 0.499, *P*=.04), and family history of dementia (OR 1.933, *P*=.04) were significantly associated with the experience of cognitive training apps for the prevention of dementia.

Conclusions: This study may serve as a theoretical basis for the development of intervention strategies to increase the use of cognitive training apps for the prevention of dementia.

(J Med Internet Res 2022;24(1):e31664) doi:10.2196/31664

KEYWORDS

cognitive training apps; dementia knowledge; health belief model; middle-aged; logistic regression analysis; dementia; Alzheimer disease; cognition; mobile apps; health apps

Introduction

Background

Population aging is a worldwide phenomenon, and the proportion of the elderly is increasing dramatically in developed countries. Along with this trend, the prevalence of dementia, which mainly affects older people, is also growing substantially. Currently, 6.2 million (11.3%) Americans aged 65 years or older have Alzheimer disease (AD), which is expected to grow more than twice by 2050 [1]. Similarly, 0.75 million (10.2%) South Koreans aged 65 or older are living with dementia, which is estimated to reach 3 million (16.1%) by 2050 [2]. The cost of health care for dementia patients is also becoming a huge burden, annually spending US $355 billion in the United States [1] and US $1 billion in South Korea [2].

Although AD commonly affects the elderly, the first patient reported was a middle-aged person [3]. Early-onset AD, which refers to a diagnosis of AD before the age of 65 years, accounts for about 5% of all AD [4]. The main symptoms may seem similar, although the diagnosis is greatly delayed and shows a more aggressive course of disease [5]. General management strategies for early-onset AD are similar to senile AD, but targeted cognitive therapies and age-appropriate psychosocial support are critical [5]. Awareness of the disease entity is relatively low, and most studies on AD are focused on adults over 65 years of age, leaving out the early onset population. The middle-aged are a unique age group. There are many middle-aged people who, on the one hand, experience with parents or relatives suffering from dementia; on the other hand, many of them are susceptible to early-onset AD. Therefore, this group deserves more attention.

Numerous preventive measures may help manage the growing burden of AD. Meta-analysis suggested that if some modifiable risk factors for AD were reduced by 10-20% per decade, the prevalence of dementia in 2050 could be decreased by 8-15% [6]. Moreover, engaging in mentally stimulating activities can greatly decrease the risk of developing AD [7], and low education attainment is known as the largest modifiable risk factor for AD development [8]. Therefore, early intervention of risk factors and active prevention measures may greatly decrease the socioeconomic burden of dementia.

Behavioral changes leading to the prevention of dementia are greatly influenced by personal beliefs that dementia can be prevented or by personal experiences with patients with dementia [9]. These perceptions about dementia can also be negatively affected by inaccurate knowledge about dementia. If someone's level of awareness for the causes and symptoms of dementia is low, they are less likely to participate in preventive activities for dementia, which may lead to delayed diagnosis and treatment [10,11]. Therefore, knowledge and perceptions about dementia is very important for the proper management of the disease [12]. However, according to a systematic review on the knowledge about dementia in the general population, 19 out of 40 reports showed that the level of dementia knowledge was low or very limited [13]. Another study found that the middle-aged population showed low levels of knowledge regarding dementia risk factors [14,15]. Therefore, the middle-aged may perceive dementia as a nonpreventable disease and may be less likely to take preventive measures.

Today, various mobile apps are being developed and utilized for the early detection or prevention of health conditions. Apps developed for dementia prevention mainly focus on cognitive training or stimulation, which may improve cognitive functions such as memory, concentration, or visuospatial coordination [16-20]. Many studies have shown that using these apps can improve memory and enhance quality of life, both in healthy adults and individuals with mild cognitive impairment [20-24].

Despite the reported usefulness of this technology, factors associated with the use of cognitive training apps for dementia prevention have not been reported in the literature. The use of cognitive training apps for dementia prevention is a combination of health-related behavior and the acceptance of technology. The health belief model (HBM) is widely used to predict the determinants related to health-related behaviors, which evaluates constructs including perceived benefits, perceived barriers, perceived susceptibility, and perceived severity to understand the likelihood of behavior [25]. The technology acceptance model is used to evaluate the causes that affect people to accept or reject technology, which measures perceived ease of use and perceived usefulness to explain usage intentions and behavior [26,27]. As each model has its limitations to explain the use of health-related apps, many recent studies have attempted to extend or combine models for better explanations of the factors that affect the acceptance of mobile health care apps [28-31].

Objective

The aim of this study was to investigate the association of structural variables and perceived health belief constructs with the experience of cognitive training apps in middle-aged adults, based on the HBM. Perceived benefits and barriers were measured for the use of cognitive training apps, and perceived susceptibility and severity was measured for dementia. Additionally, we included dementia knowledge as an additional variable to increase the explanatory power of our model.

Methods

Recruitment

Data were collected from middle-aged adults in South Korea aged 40 to 64 years between February 4 and February 8, 2021. The participants were recruited online by a professional agency (Macromill Embrain Co), where about 1.3 million participants from the general population are maintained by a demographic distribution based on census data from the National Statistical Office. First, a weblink or notice was sent to the participants via email or mobile app, whereby all participants were informed about the purpose of the study via online documentation on the starting page. The information was available for download if needed. The participants voluntarily moved onto the survey by clicking the start button, which was considered as an informed consent. Next, the participants were asked to respond to an online questionnaire. The full survey is provided in Multimedia Appendices 1 and 2. One of the unique features of an online survey is that the survey does not progress to the next question if an answer is omitted or inaccurate, which prevents incomplete or inaccurate data. The participants were free to drop out at any point of the survey if they wanted to. All personal identifying information was removed from the collected data. This study was approved by the institutional review board of Seoul National University (2102/002-002).

A total of 547 participants initially accessed the survey; 2 people did not satisfy the age criteria, and 39 dropped out before completion. Of the remaining 506 participants, 362 were selected using the proportionate quota sampling method for age and gender, due to oversampling. Participants in their 40s, 50s, and 60s were included at a 2:2:1 ratio, and men and women were selected at a 1:1 ratio, respectively. Of the selected 362 people, 42 were excluded due to poor data quality, which included extremely short response times or giving same answers for all items. In conclusion, 320 participants were included in the final analysis.

Measures

The participants rated each of the following items on a 5-point Likert scale (1=strongly disagree to 5=strongly agree), unless stated otherwise.

General Characteristics

Age, gender, education level, marital status, chronic diseases, family history of dementia, and experience of using cognitive training apps were assessed with standard survey items.

Dementia Knowledge

Dementia knowledge was measured with a dementia awareness scale developed for a national survey on the prevalence of dementia in South Korea [2], which evaluates the individual's knowledge about various aspects of dementia. The scale consists of 15 items which are answered "yes," "no," or "don't know" (to account for false positives), and the number of correct answers is summed up to a final score.

Perceived Benefit and Barrier of Using Cognitive Training Apps

The perceived benefit of using apps for cognitive training was measured with 4 items adapted from Venkatesh and Davis [27]. The questions were originally used to measure the perceived usefulness of accepting technology, which were modified for using apps and adapted in Korean for this study. The perceived barrier of using apps for cognitive training was measured with a modified set of 5 items described previously [32]. The items were originally used to measure the perceived barriers of using mobile health apps, which were modified for cognitive training apps in this study.

Perceived Susceptibility and Severity of Dementia

The perceived susceptibility of dementia and severity of dementia was measured with 4 items each, derived from the intention-to-screen questionnaire originally described by Galvin et al [33] and adapted in Korean by Yoo and Kim [34].

Statistical Analysis

Data were analyzed with SPSS, version 22.0 (IBM Corp). The differences between the groups were analyzed using the Student *t* test or one-way analysis of variance (ANOVA), and $P<.05$ was considered statistically significant. We performed logistic regression analysis to examine the relationships between the measured variables and the experience of cognitive training apps.

Results

General Characteristics

The general characteristics of the study sample are shown in Table 1. Of the 320 respondents, 82 (25.6%) had experience with cognitive training apps, while 238 (74.4%) did not. We compared the general characteristics between these 2 groups. Within the study sample, 62.2% (n=51) of participants from the experienced group were female, compared to 47.1% (n=112) for the nonexperienced group ($X^2=5.591$, $P=.02$). Participants with a family history of dementia accounted for 29.3% (n=24) in the experienced group, which was significantly higher compared with 16.8% (n=40) for the nonexperienced group ($X^2=5.919$, $P=.02$).

Table 1. Comparison of general characteristics according to experience of cognitive training apps (N=320).

Variables	All participants	Experience of cognitive training apps			
	(N=320)	Yes (n=82)	No (n=238)	X^2	*P* value
Age range (years), n (%)				2.850	.24
40-49	124 (38.8)	26 (31.7)	98 (41.2)		
50-59	129 (40.3)	39 (47.6)	90 (37.8)		
60-64	67 (20.9)	17 (20.7)	50 (21.0)		
Gender, n (%)				5.591	.02
Male	157 (49.1)	31 (37.8)	126 (52.9)		
Female	163 (50.9)	51 (62.2)	112 (47.1)		
Education level, n (%)				2.506	.29
High school	74 (23.1)	23 (28.0)	51 (21.4)		
College	211 (65.9)	53 (64.6)	158 (66.4)		
Graduate school	35 (10.9)	6 (7.3)	29 (12.2)		
Marital status, n (%)				0.366	.83
Single	40 (12.5)	9 (11.0)	31 (13.0)		
Married	256 (80.0)	65 (80.5)	190 (79.8)		
Other	24 (7.5)	7 (8.5)	17 (7.1)		
Chronic diseases, n (%)				0.381	.54
Yes	112 (35.0)	31 (37.8)	81 (34.0)		
No	208 (65.0)	51 (62.2)	157 (66.0)		
Family history of dementia, n (%)				5.919	.02
Yes	64 (20.0)	24 (29.3)	40 (16.8)		
No or other	256 (80.0)	58 (70.7)	198 (83.2)		

Descriptive Statistics of Study Variables

Descriptive statistics of the measured study variables according to experience of cognitive training apps are shown in Table 2. The participants with experience of cognitive training apps showed higher levels of dementia knowledge (*P*<.001) and perceived benefit of using cognitive training apps (*P*<.001), compared with nonexperienced individuals. The perceived barrier of using cognitive training apps was lower (*P*=.02) in the experienced group. Both perceived susceptibility and severity of dementia did not show significant differences between the groups.

Table 2. Comparison of measured variables according to experience of cognitive training apps (N=320).

Variables	All participants	Experience of cognitive training apps			
	(N=320)	Yes (n=82)	No (n=238)	*t* value	*P* value
Dementia knowledge, mean (SD)	9.05 (2.28)	9.87 (2.42)	8.77 (2.17)	3.817	<.001
Perceived benefit of using apps, mean (SD)	14.92 (2.47)	16.16 (2.12)	14.50 (2.45)	5.875	<.001
Perceived barrier of using apps, mean (SD)	13.44 (3.08)	12.76 (3.27)	13.68 (2.98)	-2.354	.02
Perceived susceptibility of dementia, mean (SD)	11.45 (2.83)	11.04 (3.05)	11.59 (2.75)	-1.524	.13
Perceived severity of dementia, mean (SD)	14.51 (2.40)	14.62 (2.54)	14.47 (2.35)	.506	.61

Factors Associated With the Experience of Cognitive Training Apps

Based on the statistical analyses above, we included 2 general characteristics (gender and family history of dementia) and 3 measured variables (dementia knowledge, perceived benefit, and perceived barrier) as possible predicting factors of the experience of cognitive training apps for logistic regression analysis (Table 3). The results revealed that higher dementia knowledge (odds ratio [OR] 1.164, *P*=.02), higher perceived benefit (OR 1.373, *P*<.001), female gender (OR 0.548, *P*=.04), and family history of dementia (OR 1.933, *P*=.04) showed positive relationships with experience of cognitive training apps for the prevention of dementia.

Table 3. Predicting factors of experience of cognitive training apps.

Variables	B[a]	SE	OR[b]	95% CI	*P* value
Dementia knowledge	0.152	0.067	1.164	1.021-1.328	.02
Perceived benefit of using apps	0.317	0.072	1.373	1.192-1.581	<.001
Perceived barrier of using apps	-0.033	0.047	0.499	0.480-0.967	.48
Gender (male=1)	-0.601	0.285	.548	0.314-0.958	.04
Family history of dementia (yes=1)	0.659	0.327	1.933	1.018-3.669	.04
Constant	-8.314	1.882	0.000	—[c]	—

[a]B: unstandardized regression weight.

[b]OR: odds ratio.

[c]Not applicable.

Discussion

Principal Findings

This study investigated the factors associated with the actual use of cognitive training apps for the prevention of dementia in middle-aged adults. Among the 320 participants, only 82 (25.6%) reported to have experience with cognitive training apps, which is still quite low, considering the widespread distribution and frequent usage of smartphones and related apps in South Korea. As the amount of evidence on the efficacy of cognitive training apps for improving cognitive functions or preventing dementia is increasing [21-23], facilitating the use of mobile apps for dementia care could be an easy solution with multiple positive effects. Conventional offline dementia prevention programs require physical space and human resources, and physical disabilities of an individual may limit access to the program [35]. Innovative programs that utilize mobile apps can overcome these hurdles, which may also lead to continued use. In addition, gamified cognitive training apps are fun and motivational, cost very little, and can be performed in a comfortable environment at a convenient time [36], which could greatly increase adherence. However, one must note that not all cognitive training apps are based on scientific evidence, and therefore these ideas should not be generalized. Further studies investigating the effects of specific cognitive training apps and the types of cognitive training included are needed, which would shed light on how to select a cognitive training app for the prevention of dementia.

Our results showed that the perceived benefit of using cognitive training apps was positively associated with the use of cognitive training apps, but other HBM constructs did not. Previous studies also suggest that perceived benefit is a consistent predictor of health-related behavior, while perceived susceptibility and severity often failed to explain health-related behavior [37-39]. Additionally, previous studies using the technology acceptance model also showed that the perceived usefulness of using apps is a positive predictor of the intention to use the apps [40-42]. In other words, believing that using cognitive training apps can improve cognitive function is an important factor in predicting the intention to use the apps. Moreover, this belief is also critical for the continued use of health apps [43,44].

Dementia knowledge was linked to the actual use of cognitive training apps in this study, which is consistent with previous reports showing that dementia knowledge is positively associated with preventive behavior for dementia [45-47]. A systematic review on dementia knowledge points out that most people think cognitive activities that exercise the brain are more effective in preventing dementia rather than medications, exercise, or dietary modifications [48], which may explain the link between dementia knowledge and the use of cognitive training apps.

Meanwhile, the average dementia knowledge score in this study was 9.05 points out of 15. Recent studies on middle-aged adults (40 to 75 years of age) showed that more than half of the respondents had insufficient dementia knowledge, 59.4% wanted information for cognitive health, and 70% had positive feelings for eHealth use to improve cognitive health [14,15]. As both the need and demand for dementia-related education is high, mobile apps can serve as a useful tool to deliver dementia knowledge and provide cognitive training programs in the middle-aged compared to the elderly, in terms of technology friendliness.

Our results showed that women were more likely to use cognitive training apps than men. This is consistent with a previous study, which showed women participate in activities that help improve cognitive health more frequently than men [49]. This may be linked to the higher prevalence of dementia among women compared with men [50], and the higher level of experience with dementia patients in females [1], although further studies are required to elucidate the exact relationships between these factors.

Finally, individuals with a family history of dementia were more likely to use cognitive training apps in this study. Previous studies show that people with a family history of dementia think that they have a higher risk of developing dementia, and as having a family history of AD has consistently emerged as a key predictor of dementia worry [51], they may undergo activities for dementia prevention more actively. However, our results showed that perceived susceptibility did not predict the use of cognitive training apps, similar to the findings of a previous report that showed self-perceived risk itself did not predict preventive behavior [52]. Other studies also show that people with family history of dementia are less likely to believe

that dementia is preventable, and that they have lower self-efficacy for dementia prevention [53]. Therefore, it is important to understand the characteristics of this population and perform suitable interventions, which can lead to a positive attitude for dementia prevention.

Limitations

The questionnaire data was collected on an online basis; therefore, individuals could have shown higher digital literacy compared with the normal population.

Conclusions

This study explores the influencing factors on the experience of cognitive training apps using an extended HBM model, which may serve as a theoretical basis for the development of intervention strategies to increase the use of cognitive training apps for dementia prevention.

Authors' Contributions

JL and JML conceived and designed the study, analyzed data, and wrote the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Full questionnaire (in Korean).
[DOCX File , 766 KB - jmir_v24i1e31664_app1.docx]

Multimedia Appendix 2
Full questionnaire (in English).
[DOCX File , 765 KB - jmir_v24i1e31664_app2.docx]

References

1. Alzheimer's Association. 2021 Alzheimer's disease facts and figures. Alzheimers Dement 2021 Mar;17(3):327-406. [doi: 10.1002/alz.12328] [Medline: 33756057]
2. Korean Dementia Observatory 2019. National Institute of Dementia.: National Institute of Dementia; 2020 Feb 28. URL: https://www.nid.or.kr/info/dataroom_view.aspx?bid=209 [accessed 2021-08-24]
3. Maurer K, Volk S, Gerbaldo H. Auguste D and Alzheimer's disease. Lancet 1997 May 24;349(9064):1546-1549. [doi: 10.1016/S0140-6736(96)10203-8] [Medline: 9167474]
4. Zhu X, Tan L, Wang H, Jiang T, Cao L, Wang C, et al. Rate of early onset Alzheimer's disease: a systematic review and meta-analysis. Ann Transl Med 2015 Mar;3(3):38. [doi: 10.3978/j.issn.2305-5839.2015.01.19] [Medline: 25815299]
5. Mendez MF. Early-Onset Alzheimer Disease. Neurol Clin 2017 May;35(2):263-281 [FREE Full text] [doi: 10.1016/j.ncl.2017.01.005] [Medline: 28410659]
6. Norton S, Matthews FE, Barnes DE, Yaffe K, Brayne C. Potential for primary prevention of Alzheimer's disease: an analysis of population-based data. Lancet Neurol 2014 Aug;13(8):788-794. [doi: 10.1016/S1474-4422(14)70136-X] [Medline: 25030513]
7. Crous-Bou M, Minguillón C, Gramunt N, Molinuevo JL. Alzheimer's disease prevention: from risk factors to early intervention. Alzheimers Res Ther 2017 Sep 12;9(1):71 [FREE Full text] [doi: 10.1186/s13195-017-0297-z] [Medline: 28899416]
8. Barnes DE, Yaffe K. The projected effect of risk factor reduction on Alzheimer's disease prevalence. Lancet Neurol 2011 Sep;10(9):819-828 [FREE Full text] [doi: 10.1016/S1474-4422(11)70072-2] [Medline: 21775213]
9. Low L, Anstey KJ. The public's perception of the plausibility of dementia risk factors is not influenced by scientific evidence. Dement Geriatr Cogn Disord 2007;23(3):202-206. [doi: 10.1159/000099038] [Medline: 17259711]
10. Akyol MA, Zehirlioğlu L, Erünal M, Mert H, Hatipoğlu NS, Küçükgüçlü Ö. Determining Middle-Aged and Older Adults' Health Beliefs to Change Lifestyle and Health Behavior for Dementia Risk Reduction. Am J Alzheimers Dis Other Demen 2020 Jan;35:1533317519898996 [FREE Full text] [doi: 10.1177/1533317519898996] [Medline: 32048860]
11. Leung AYM, Molassiotis A, Zhang J, Deng R, Liu M, Van IK, et al. Dementia Literacy in the Greater Bay Area, China: Identifying the At-Risk Population and the Preferred Types of Mass Media for Receiving Dementia Information. Int J Environ Res Public Health 2020 Apr 07;17(7):2511 [FREE Full text] [doi: 10.3390/ijerph17072511] [Medline: 32272551]
12. Nationwide Survey on the Dementia Epidemiology of Korea. National Institute of Dementia. 2017 Jun 30. URL: https://www.nid.or.kr/info/dataroom_view.aspx?bid=182 [accessed 2021-06-10]
13. Cahill S, Pierce M, Werner P, Darley A, Bobersky A. A systematic review of the public's knowledge and understanding of Alzheimer's disease and dementia. Alzheimer Dis Assoc Disord 2015;29(3):255-275. [doi: 10.1097/WAD.0000000000000102] [Medline: 26207322]

14. Van Asbroeck S, van Boxtel MPJ, Steyaert J, Köhler S, Heger I, de Vugt M, et al. Increasing knowledge on dementia risk reduction in the general population: Results of a public awareness campaign. Prev Med 2021 Jun;147:106522 [FREE Full text] [doi: 10.1016/j.ypmed.2021.106522] [Medline: 33744328]
15. Heger I, Köhler S, van Boxtel M, de Vugt M, Hajema K, Verhey F, et al. Raising awareness for dementia risk reduction through a public health campaign: a pre-post study. BMJ Open 2020 Nov 06;10(11):e041211 [FREE Full text] [doi: 10.1136/bmjopen-2020-041211] [Medline: 33158836]
16. Hill NL, Mogle J, Wion R, Kitt-Lewis E, Hannan J, Dick R, et al. App-based attention training: Incorporating older adults' feedback to facilitate home-based use. Int J Older People Nurs 2018 Mar;13(1):e12163. [doi: 10.1111/opn.12163] [Medline: 28752612]
17. Lu M, Lin W, Yueh H. Development and Evaluation of a Cognitive Training Game for Older People: A Design-based Approach. Front Psychol 2017 Oct 17;8:1837. [doi: 10.3389/fpsyg.2017.01837] [Medline: 29089914]
18. Powell LE, Wild MR, Glang A, Ibarra S, Gau JM, Perez A, et al. The development and evaluation of a web-based programme to support problem-solving skills following brain injury. Disabil Rehabil Assist Technol 2019 Jan;14(1):21-32. [doi: 10.1080/17483107.2017.1389999] [Medline: 29063800]
19. Tacchino A, Pedullà L, Bonzano L, Vassallo C, Battaglia MA, Mancardi G, et al. A New App for At-Home Cognitive Training: Description and Pilot Testing on Patients with Multiple Sclerosis. JMIR Mhealth Uhealth 2015 Aug 31;3(3):e85 [FREE Full text] [doi: 10.2196/mhealth.4269] [Medline: 26323749]
20. Oh SJ, Seo S, Lee JH, Song MJ, Shin M. Effects of smartphone-based memory training for older adults with subjective memory complaints: a randomized controlled trial. Aging Ment Health 2018 Apr;22(4):526-534. [doi: 10.1080/13607863.2016.1274373] [Medline: 28071929]
21. Yasini M, Marchand G. Adoption and Use of a Mobile Health Application in Older Adults for Cognitive Stimulation. Stud Health Technol Inform 2016;221:13-17. [doi: 10.3233/978-1-61499-633-0-13] [Medline: 27071867]
22. Shellington EM, Felfeli T, Shigematsu R, Gill DP, Petrella RJ. HealthBrain: an innovative smartphone application to improve cognitive function in older adults. Mhealth 2017 May;3(5):17 [FREE Full text] [doi: 10.21037/mhealth.2017.04.05] [Medline: 28567413]
23. Han JW, Son KL, Byun HJ, Ko JW, Kim K, Hong JW, et al. Efficacy of the Ubiquitous Spaced Retrieval-based Memory Advancement and Rehabilitation Training (USMART) program among patients with mild cognitive impairment: a randomized controlled crossover trial. Alzheimers Res Ther 2017 Jun 06;9(1):39 [FREE Full text] [doi: 10.1186/s13195-017-0264-8] [Medline: 28587629]
24. Martínez-Alcalá CI, Rosales-Lagarde A, Hernández-Alonso E, Melchor-Agustin R, Rodriguez-Torres EE, Itzá-Ortiz BA. A Mobile App (iBeni) With a Neuropsychological Basis for Cognitive Stimulation for Elderly Adults: Pilot and Validation Study. JMIR Res Protoc 2018 Aug 21;7(8):e172 [FREE Full text] [doi: 10.2196/resprot.9603] [Medline: 30131319]
25. Rosenstock IM. Historical Origins of the Health Belief Model. Health Education Monographs 1974 Dec 01;2(4):328-335. [doi: 10.1177/109019817400200403]
26. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
27. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
28. Ahadzadeh AS, Pahlevan Sharif S, Ong FS, Khong KW. Integrating health belief model and technology acceptance model: an investigation of health-related internet use. J Med Internet Res 2015 Feb 19;17(2):e45 [FREE Full text] [doi: 10.2196/jmir.3564] [Medline: 25700481]
29. Bae H, Jo SH, Lee E. Influence of Locus of Control and Belief in Health Promotion Behavior on the Intention to Use Mobile Healthcare Services. Journal of Marketing Management Research 2019 Apr;24(2):55-71. [doi: 10.37202/KMMR.2019.24.2.55]
30. Moon J, Rigg JS, Smith JE. Korean American Smokers' Perspectives on Mobile Smoking Cessation Applications. Tob Use Insights 2020 Dec 13;13:1179173X20972384 [FREE Full text] [doi: 10.1177/1179173X20972384] [Medline: 33402856]
31. Hong W, Liu R, Ding Y, Hwang J, Wang J, Yang Y. Cross-Country Differences in Stay-at-Home Behaviors during Peaks in the COVID-19 Pandemic in China and the United States: The Roles of Health Beliefs and Behavioral Intention. Int J Environ Res Public Health 2021 Feb 21;18(4):2104 [FREE Full text] [doi: 10.3390/ijerph18042104] [Medline: 33670078]
32. Wang JS. A Study on Factors Influencing the Individual's Health Beliefs on the Intention of Mobile Health Care Applications: Focusing on the e-health literacy control effects (MA Thesis). Research Information Sharing Service.: Kyung Hee University; 2020. URL: https://tinyurl.com/2s35bc48 [accessed 2021-06-30]
33. Galvin JE, Scharff DP, Glasheen C, Fu Q. Development of a population-based questionnaire to explore psychosocial determinants of screening for memory loss and Alzheimer Disease. Alzheimer Dis Assoc Disord 2006;20(3):182-191. [doi: 10.1097/00002093-200607000-00010] [Medline: 16917189]
34. Yoo R, Kim GS. Factors Affecting the Performance of the Dementia Screening Test Using the Health Belief Model. Journal of Korean Public Health Nursing 2017 Dec;31(3):464-477 Print ISSN: 1226-0290 / Online ISSN: 2234-2869. [doi: 10.5932/JKPHN.2017.31.3.464]

35. Kim SJ, Choi YJ. Predictable Effect and Usability of Smart-Phone Application for Elderly Dementia Prevention. Journal of Industrial Convergence 2019 Dec 31;17(4):87-94. [doi: 10.22678/jic.2019.17.4.087]
36. Giannopoulou P, Vlamos P. Analysis and design of an information system for cognitive training of patients with mild cognitive impairment using mobile devices. 2020 Sep Presented at: 2020 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM); 2020 Sep 25-27; Virtual Event. [doi: 10.1109/seeda-cecnsm49515.2020.9221782]
37. Floyd DL, Prentice-Dunn S, Rogers RW. A Meta-Analysis of Research on Protection Motivation Theory. J Appl Social Pyschol 2000 Feb;30(2):407-429. [doi: 10.1111/j.1559-1816.2000.tb02323.x]
38. Milne S, Sheeran P, Orbell S. Prediction and Intervention in Health-Related Behavior: A Meta-Analytic Review of Protection Motivation Theory. J Appl Social Pyschol 2000 Jan;30(1):106-143. [doi: 10.1111/j.1559-1816.2000.tb02308.x]
39. Jones CJ, Smith H, Llewellyn C. Evaluating the effectiveness of health belief model interventions in improving adherence: a systematic review. Health Psychol Rev 2014;8(3):253-269. [doi: 10.1080/17437199.2013.802623] [Medline: 25053213]
40. Jo SC, Han YJ. A Study on the Effect of Health Belief Factors on the Acceptance of Mobile Healthcare : Focusing on Mediating Effects of Perceived Usefulness. RIR 2020 May 30;43(2):263-280. [doi: 10.33932/rir.43.2.12]
41. Yee TS, Seong LC, Chin WS. Patient's Intention to Use Mobile Health App. Journal of Management Research 2019 May 28;11(3):18-35 [FREE Full text] [doi: 10.5296/jmr.v11i3.14776]
42. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use Medical Apps Among Older Adults in the Netherlands: Cross-Sectional Study. J Med Internet Res 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: 10.2196/18080] [Medline: 32624465]
43. Bhattacherjee A. Understanding Information Systems Continuance: An Expectation-Confirmation Model. MIS Quarterly 2001 Sep;25(3):351-370 [FREE Full text] [doi: 10.2307/3250921]
44. Cho J. The impact of post-adoption beliefs on the continued use of health apps. Int J Med Inform 2016 Mar;87:75-83. [doi: 10.1016/j.ijmedinf.2015.12.016] [Medline: 26806714]
45. Devoy S, Simpson EEA. Help-seeking intentions for early dementia diagnosis in a sample of Irish adults. Aging Ment Health 2017 Aug;21(8):870-878. [doi: 10.1080/13607863.2016.1179262] [Medline: 27149181]
46. Kim S, Sargent-Cox KA, Anstey KJ. A qualitative study of older and middle-aged adults' perception and attitudes towards dementia and dementia risk reduction. J Adv Nurs 2015 Jul;71(7):1694-1703. [doi: 10.1111/jan.12641] [Medline: 25702823]
47. Park J, Shin JH. Factors Influencing the Dementia-Preventive Behaviors among Middle-Aged Persons with Chronic Diseases in Korea. Int J Environ Res Public Health 2021 Mar 13;18(6):2936 [FREE Full text] [doi: 10.3390/ijerph18062936] [Medline: 33805592]
48. Cations M, Radisic G, Crotty M, Laver KE. What does the general public understand about prevention and treatment of dementia? A systematic review of population-based surveys. PLoS One 2018 Apr 19;13(4):e0196085 [FREE Full text] [doi: 10.1371/journal.pone.0196085] [Medline: 29672559]
49. Smith BJ, Ali S, Quach H. The motivation and actions of Australians concerning brain health and dementia risk reduction. Health Promot J Austr 2015 Aug;26(2):115-121. [doi: 10.1071/HE14111] [Medline: 26210295]
50. Werner P, Goldberg S, Mandel S, Korczyn AD. Gender differences in lay persons' beliefs and knowledge about Alzheimer's disease (AD): a national representative study of Israeli adults. Arch Gerontol Geriatr 2013;56(2):400-404. [doi: 10.1016/j.archger.2012.11.001] [Medline: 23219063]
51. Kessler E, Bowen CE, Baer M, Froelich L, Wahl H. Dementia worry: a psychological examination of an unexplored phenomenon. Eur J Ageing 2012 Dec;9(4):275-284 [FREE Full text] [doi: 10.1007/s10433-012-0242-8] [Medline: 28804427]
52. Maxfield M, Greenberg J. Anticipated Stigma and Dementia-Related Anxiety in Middle-Aged and Older Adults. GeroPsych (Bern) 2021 Mar;34(1):13-22. [doi: 10.1024/1662-9647/a000234] [Medline: 34276334]
53. Lee W, Gray SL, Zaslavsky O, Barthold D, Marcum ZA. Association between having a family member with dementia and perceptions of dementia preventability. Aging Ment Health 2020 Nov 02:1-7. [doi: 10.1080/13607863.2020.1839866] [Medline: 33131283]

Abbreviations

AD: Alzheimer disease
ANOVA: analysis of variance
HBM: health belief model
OR: odds ratio

Edited by R Kukafka; submitted 30.06.21; peer-reviewed by G Shakerinejad, R Eckhoff; comments to author 13.11.21; revised version received 22.11.21; accepted 03.12.21; published 14.01.22.

Please cite as:
Lee J, Lim JM
Factors Associated With the Experience of Cognitive Training Apps for the Prevention of Dementia: Cross-sectional Study Using an Extended Health Belief Model
J Med Internet Res 2022;24(1):e31664
URL: https://www.jmir.org/2022/1/e31664
doi:10.2196/31664
PMID:35029540

Original Paper

Engagement With a Mobile Phone–Based Life Skills Intervention for Adolescents and Its Association With Participant Characteristics and Outcomes: Tree-Based Analysis

Raquel Paz Castro[1], PhD; Severin Haug[1], PhD; Rudolf Debelak[2,3], PhD; Robert Jakob[4], MSc; Tobias Kowatsch[4,5], PhD; Michael P Schaub[1], PhD

[1]Swiss Research Institute for Public Health and Addiction, University of Zurich, Zurich, Switzerland

[2]Department of Psychology, Psychological Methods, Evaluation and Statistics, University of Zurich, Zurich, Switzerland

[3]Wilhelm Wundt Institute for Psychology, University of Leipzig, Leipzig, Germany

[4]Centre for Digital Health Interventions, Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

[5]Centre for Digital Health Interventions, Institute of Technology Management, University of St.Gallen, St.Gallen, Switzerland

Corresponding Author:
Raquel Paz Castro, PhD
Swiss Research Institute for Public Health and Addiction
University of Zurich
Konradstrasse 32
Zurich, 8051
Switzerland
Phone: 41 444481180
Email: raquel.paz@isgf.uzh.ch

Abstract

Background: Mobile phone–delivered life skills programs are an emerging and promising way to promote mental health and prevent substance use among adolescents, but little is known about how adolescents actually use them.

Objective: The aim of this study is to determine engagement with a mobile phone–based life skills program and its different components, as well as the associations of engagement with adolescent characteristics and intended substance use and mental health outcomes.

Methods: We performed secondary data analysis on data from the intervention group (n=750) from a study that compared a mobile phone–based life skills intervention for adolescents recruited in secondary and upper secondary school classes with an assessment-only control group. Throughout the 6-month intervention, participants received 1 SMS text message prompt per week that introduced a life skills topic or encouraged participation in a quiz or individual life skills training or stimulated sharing messages with other program participants through a friendly contest. Decision trees were used to identify predictors of engagement (use and subjective experience). The stability of these decision trees was assessed using a resampling method and by graphical representation. Finally, associations between engagement and intended substance use and mental health outcomes were examined using logistic and linear regression analyses.

Results: The adolescents took part in half of the 50 interactions (mean 23.6, SD 15.9) prompted by the program, with SMS text messages being the most used and contests being the least used components. Adolescents who did not drink in a problematic manner and attended an upper secondary school were the ones to use the program the most. Regarding associations between engagement and intended outcomes, adolescents who used the contests more frequently were more likely to be nonsmokers at follow-up than those who did not (odds ratio 0.86, 95% CI 0.76-0.98; *P*=.02). In addition, adolescents who read the SMS text messages more attentively were less likely to drink in a problematic manner at follow-up (odds ratio 0.43, 95% CI 1.29-3.41; *P*=.003). Finally, participants who used the program the most and least were more likely to increase their well-being from baseline to 6-month follow-up compared with those with average engagement (βs=.39; t_{586}=2.66; *P*=.008; R^2=0.24).

Conclusions: Most of the adolescents participating in a digital life skills program that aimed to prevent substance use and promote mental health engaged with the intervention. However, measures to increase engagement in problem drinkers should be considered. Furthermore, efforts must be made to ensure that interventions are engaging and powerful across different educational levels. First results indicate that higher engagement with digital life skills programs could be associated with intended outcomes.

Future studies should apply further measures to improve the reach of lower-engaged participants at follow-up to establish such associations with certainty.

(J Med Internet Res 2022;24(1):e28638) doi:10.2196/28638

KEYWORDS

engagement; life skills; adolescents; mobile phone; machine learning; decision tree; mobile phone

Introduction

Background

Adolescence is a vulnerable period in life, during which substance use and mental disorders often first emerge and begin to develop [1]. In the age group of 11-15 years, the prevalence of lifetime and recent alcohol and tobacco use shows a sharp increase in both genders [2]. In Switzerland, for example, the lifetime prevalence of alcohol use increased from 22% among boys aged 11 years to 70% among boys aged 15 years and from 11% in girls aged 11 years to 69% in girls aged 15 years [3]. The lifetime prevalence of tobacco use increased from 6% in boys aged 11 years to 35% in boys aged 15 years and from 2% in girls aged 11 years to 30% in girls aged 15 years. Taxation, public consumption bans, advertising restrictions, and raising the minimum legal age have proven to be effective ways to prevent substance use during adolescence. However, early interventions that incorporate skills trainings have also shown promising effectiveness [4]. Evidence for this effectiveness primarily stems from life skills trainings embedded in school curricula [5-11], whereas life skills trainings to promote mental health and prevent substance use in young adolescents are only recently being adapted to digital interventions as well as being tested [12-18]. This trend is mainly fostered by the difficulties that schools encounter when trying to implement life skills trainings in their curricula [9] and the personnel and financial resources that such programs require [19]. Digital interventions have great potential to overcome these obstacles. Such interventions have a large reach at low cost and can deliver uniquely personalized content automatically, which can be accessed at any time and anywhere [4]. The main disadvantage of delivering life skills training electronically is the lack of control over adolescents' engagement with the intervention. Several reviews on digital interventions to promote mental health [20-22] or to prevent substance use [23] in young people point at the relatively low levels of user engagement. However, the most common assumption in current literature is that some form of engagement is essential for digital interventions to be effective [24-26]. Engagement, for instance, has been conceptualized in previous studies as both "(1) the extent (e.g. amount, frequency, duration, depth) of usage and (2) a subjective experience characterised by attention, interest and affect" [25]. This definition will be used in the context of this paper. From studies on adult populations, there is some evidence on factors that may influence engagement with digital behavior change interventions [25]. The conceptual framework of Perski et al [25] specifies factors that might affect engagement, including characteristics of the intervention (eg, esthetics or design, ease of use, and the inclusion of known behavior change techniques such as feedback and goal setting) or the context (eg, norms, age, education, and self-efficacy). Other factors were hypothesized to influence engagement (eg, the target behavior itself and some mechanism of action, such as accountability and motivation). Most of the current knowledge on predictors of engagement in adolescent populations comes from digital interventions devoted to physical activity [27,28]. To date, only a few studies have examined predictors of engagement with digital interventions devoted to mental health or substance use in adolescents. Furthermore, all studies examined engagement in terms of program use and not in terms of a subjective experience. In an internet-based depression prevention program directed at young adolescents from Australian public schools, low engagement with the program was predicted by being older, living in an urban area, or having lower leves of depressive symptoms or self-esteem at baseline [27]. In a mobile phone–based smoking cessation intervention targeting Swiss adolescents from secondary schools, nonengagement over time was most common among older smokers, smokers with an immigrant background, and smokers who reported nonproblematic levels of alcohol use at baseline [29]. Because of the novelty of using digital life skills interventions to promote healthy lifestyles in adolescents, it is still unclear which characteristics predict user engagement. Ascertaining the predictors in this new field can reveal who engages the most and for whom new strategies to increase program engagement have to be developed.

Besides the question of *who engages*, there is a wider debate on the associations between engagement and outcomes. Adolescents' engagement with digital interventions on mental health has been found to be positively [30], negatively [31], or not linked [27] to intended behavior changes. Furthermore, decreasing engagement with a digital intervention—but not stable engagement—over time was associated with a smoking reduction in adolescents compared with nonengagement [29]. Which association between engagement and outcomes is observed depends on the selected engagement measure (eg, number of log-ins vs completion of a specific module) and on the underlying motivation for disengagement (eg, loss of interest vs sufficient support) [24,26,32,33]. New approaches try to operationalize *effective engagement* with the intervention, defined as the degree of engagement needed to achieve intended outcomes [34]. Given the current lack of knowledge about digital life skills interventions, a comprehensive approach to the assessment of indicators of engagement is needed. Furthermore, a broad approach must be taken to assess predictors of engagement, including previous known and unknown characteristics possibly related to program use and subjective experience. As use data tend to be highly skewed [35], methods are needed that (1) can select predictors in a consistent and unbiased way and (2) overcome distributional assumptions made by standard methods such as regression analysis.

To sum up, studies on engagement need to address different methodological challenges, through (1) a clear definition of engagement, (2) a selection of appropriate engagement measures, and (3) appropriate statistical methods that can handle large amount of data with different distributional characteristics.

Objective

This study aims to use decision trees to determine predictors of engagement within a randomized controlled trial assessing the effectiveness of a novel mobile phone–based life skills intervention for secondary and upper secondary school students. The intervention addressed (1) self-management skills, (2) social skills, and (3) substance use resistance skills, which are the central competences included in widely disseminated life skills programs [7,36,37]. Decision trees allow the selection of relevant predictors associated with the outcome of interest and display them in a way that is easy to interpret. Furthermore, the relationship between engagement and changes in outcomes will be analyzed.

Methods

Participants and Procedures

Data for this study were extracted from a 2-arm, parallel-group, cluster randomized controlled trial that used school class as the randomization unit, as detailed elsewhere [14,18].

Students in secondary and upper secondary schools in the German-speaking part of Switzerland were invited by cooperating regional centers for addiction prevention to participate in a mobile phone–based program called *SmartCoach* if they (1) were aged ≥14 years, (2) could provide informed parental consent (only necessary for those aged 14 years), and (3) owned a mobile phone. The participating students were reimbursed CHF 10 (US $10.90) for participating in the 6- and 18-month follow-up assessments. To increase program engagement, a friendly competition among the participants was integrated into the program. This form of friendly competition was encouraged by answering SMS text messages, creating messages or taking pictures within contests, or assessing video and website links integrated into SMS text messages. With every interaction, the participants were able to collect credits, and the more credits the participants collected, the higher their chances of winning 1 of 10 prizes, which were part of a prize draw (10×CHF 50 [US $54.50] in cash). The participants had constant access to their current credit score on the personal profile page, which also displayed scores of the other participants in the same group (same starting date).

In the original trial, the efficacy of the life skills intervention *SmartCoach* was tested against an assessment-only control group. A total of 1473 students (mean age 15.4, SD 1.0, years; 813/1473, 55.19%, were girls) from 89 secondary and upper secondary school classes in the German-speaking part of Switzerland participated in this study. They were randomly assigned to either the intervention (750/1473, 50.92%; 44 classes) or to the assessment-only control group (723/1473, 49.08%; 45 classes). In this study, only the participants assigned to the intervention group were examined. Of the 750 participants in the intervention group, 597 (79.6%) took part in the 6-month follow-up assessment. Attrition at 6-month follow-up was significantly associated with use of the program ($t_{423.14}$=25.3; P<.001). Participants lost to follow-up interacted a mean of 6.0 (SD 8.0) times with the program. In comparison, those who were assessed interacted a mean of 28.2 (SD 14.2) times with the program. This means that 64.8% (276/426) of the students who interacted 0-30 times with the program were assessed at follow-up in comparison with 98.4% (184/187) of those with 31-40 interactions, and 100% (137/137) of those with 41-50 interactions. The original study underlined the appropriateness of the program for the target group with most of its participants (710/750, 94.7%) staying subscribed until month 6. In addition, a large proportion of the participants stated that they read the SMS text messages (551/563, 98.4%, with valid follow-up data) and rated the program as helpful (384/550, 69.8%) and individually tailored (327/550, 59.5%). Intention-to-treat comparisons at the 6-month follow-up illustrated significant differences between both groups in terms of reducing the number of drinks consumed per month (intervention group vs control group: –0.6 vs 0.7; P=.03) and the number of cigarettes smoked per month (intervention group vs control group: –1.7 vs 5.0; P=.01), as well as reported stress (P=.02). These are initial results because the analyses of the 18-month follow-up data are ongoing.

The intervention was designed with, and initiated by, the open-source behavioral intervention platform *MobileCoach* [38,39]. The original study protocol was approved by the ethics committee of the Faculty of Arts and Sciences at the University of Zurich, Switzerland (approval number 18.6.5; date of approval: June 21, 2018). The study was registered with ISRCTN (ISRCTN41347061; assigned July 21, 2018) and executed in full compliance with the Declaration of Helsinki.

Description of SmartCoach

Figures 1 and 2 depict all components of the *SmartCoach* program. The *SmartCoach* program combined (1) tailored web-based feedback to reduce the individual level of stress, which was delivered directly after completion of the baseline assessment, and (2) tailored mobile phone SMS text messages to promote self-management skills (block 1), social skills (block 2), and substance use resistance skills (block 3) over 6 months. The theoretical backgrounds of these intervention components are described elsewhere [14,18].

During the 6-month *SmartCoach* program, participants received one SMS text message prompt per week that (1) introduced 1 of the 3 life skills blocks, (2) or encouraged the subject's participation in quizzes, self-challenges, and individual stress or social skills trainings, or (3) invited participants to post pictures or messages and vote on other posts published by all intervention participants at the end of a block. The initial SMS text messages included graphical objects but did not require an answer. The prompts encouraging the subject's participation in quizzes, self-challenges, and individual stress or social skills trainings were easily answered by typing a single letter or number using the mobile phone's reply function. Answering such a prompt could trigger 1-2 further SMS text messages in the program. Some of these answers included hyperlinks to thematically relevant video clips, pictures, and related websites.

To participate in the contests, the participants had to upload a picture or post a motivational text message on a website within 2 days. This was followed by a 2-day window for voting on all posts and a presentation of the 3 contributions with the highest votes on day 5. This function was also available on the website. Table 1 displays the sequence and contents of the SMS text messages and the possible activities prompted by these SMS text messages.

Figure 1. Overview of the *SmartCoach* survey components.

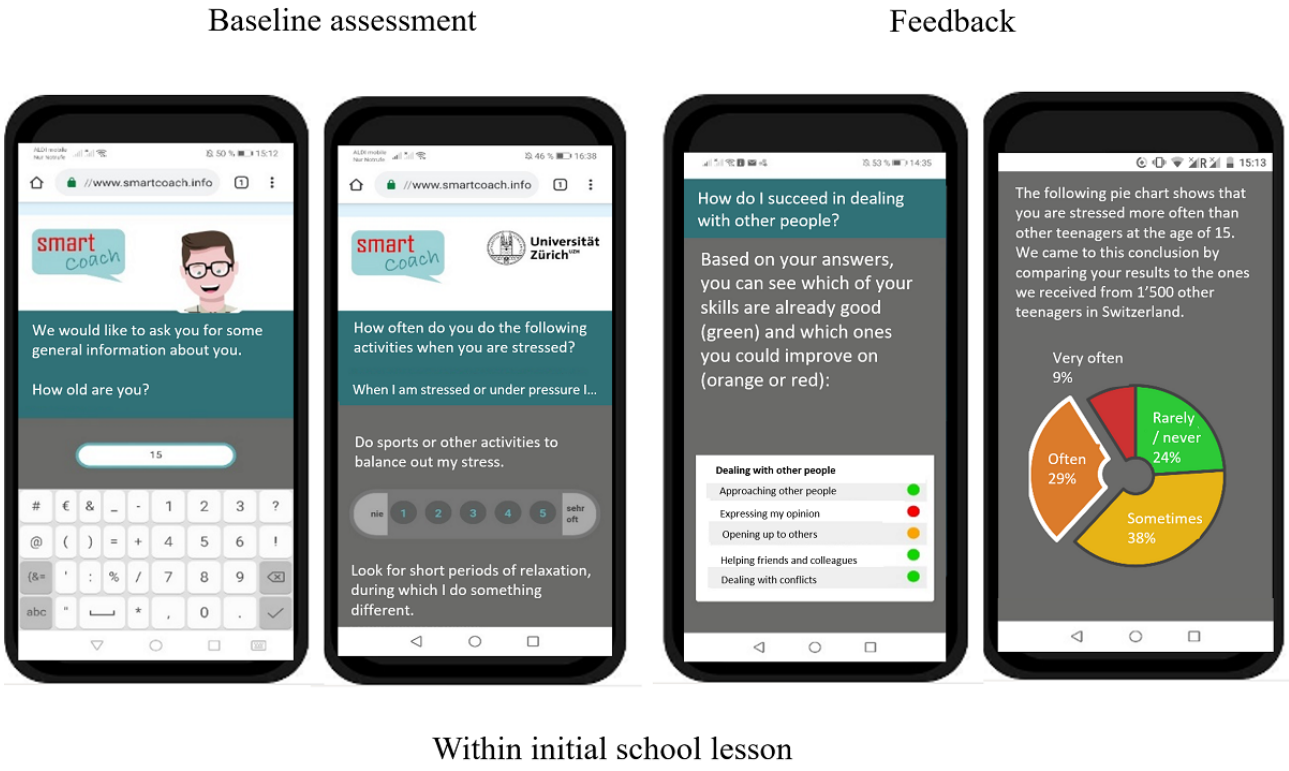


Figure 2. Overview of the *SmartCoach* intervention components.

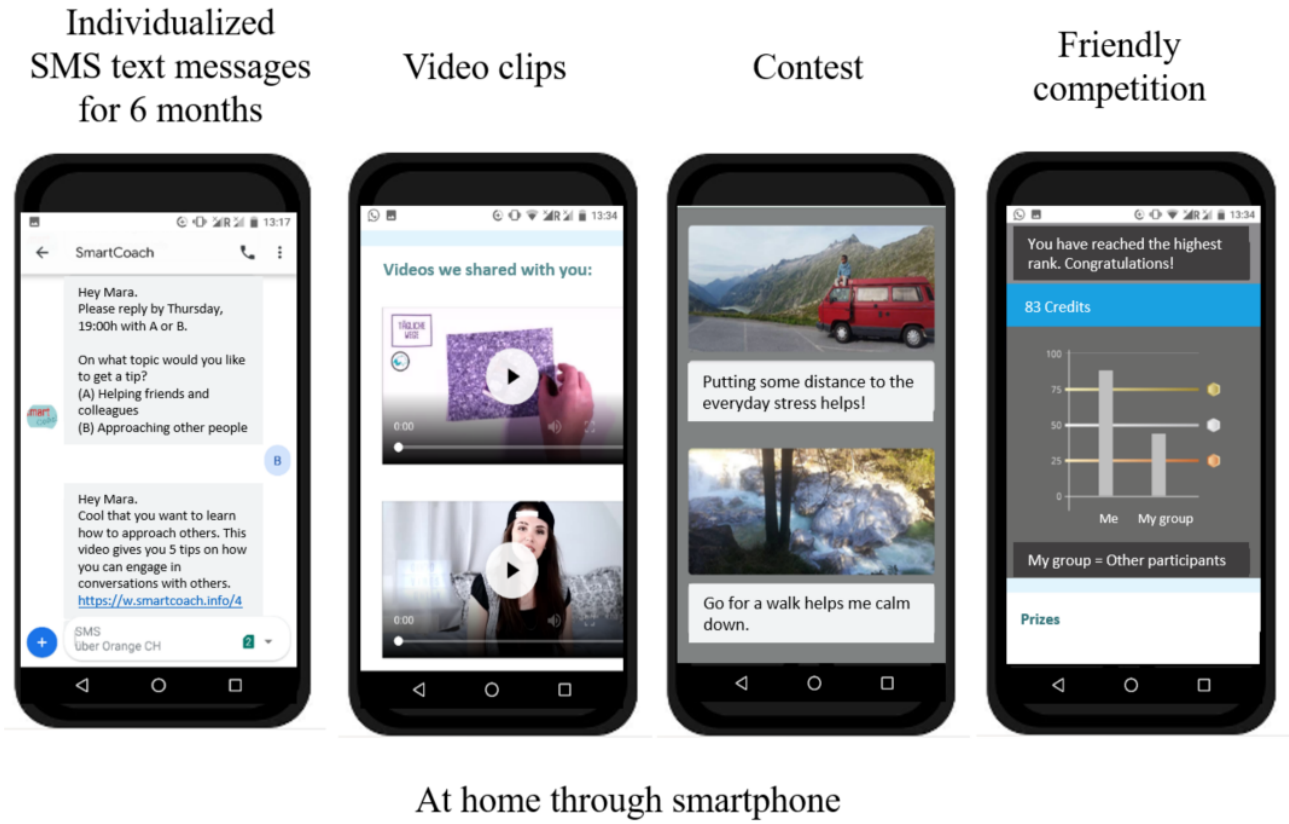


Table 1. Sequence, content, and engagement with different prompts within the *SmartCoach* program (N=750).

Week and content	Required activities	Students engaging with activity, n (%)
1: Introduction to self-management skills and quiz on origin and function of stress		
	Click on link to an overview picture	194 (25.9)
	Reply to quiz question	525 (70)
	Click on video link	444 (59.2)
2: Quiz on common stressors		
	Reply to quiz question	513 (68.4)
	Click on video link	410 (54.7)
3 and 6: Tailored stress reduction strategies for individual stressors		
	Reply to SMS text message with options	938 (62.5)[a]
	Click on video or website link	605 (40.3)[a]
4: Self-challenge on general stress reduction strategies		
	Reply to SMS text message with options	465 (62)
	Reply to SMS text message on successful application of chosen strategy	471 (62.8)
5: Quiz on eustress vs distress		
	Reply to quiz question	473 (63.1)
	Click on video link	359 (47.9)
7: Group contest on preferred stress management strategy		
	Viewing of others' posts	480 (64)[b]
	Posting a picture and message on individually preferred strategy	193 (25.7)
	Voting on others' posts	222 (29.6)
8: Introduction to social skills and quiz on social skills		
	Click on link to an overview picture	372 (49.6)
	Reply to quiz question	416 (55.5)
	Click on link to picture	301 (40.1)
9: Tailored strategies for improving personal social skills		
	Reply to SMS text message with options	716 (95.5)
	Click on video or website link	398 (53.1)
10: Quiz on use of body language in different situations		
	Reply to quiz question	426 (56.8)
	Click on video link	299 (39.9)[c]
11: Tailored strategies for improving personal social skills		
	Reply to SMS text message with options	417 (55.6)
	Click on video or website link	231 (30.9)
12: Self-challenge on strategies to improve social skills in different areas		
	Reply to SMS text message with options	403 (53.7)
	Reply to SMS text message on successful application of chosen strategy	416 (55.5)
13: Origin of smartphone addiction		
	Reply to quiz question	418 (55.7)
	Click on video link	294 (39.2)

Week and content	Required activities	Students engaging with activity, n (%)
14: Quiz on associations between smartphone use and stress tailored to gender		
	Reply to quiz question	384 (51.2)
	Click on video links	360 (48)
15: Self-challenge on smartphone detox		
	Reply to SMS text message with options	386 (51.5)
	Reply to SMS text message on successful detox in chosen situation	405 (54)
	Click on video link	239 (31.9)
16: Quiz on recognition of peer pressure		
	Click on link to the first part of the video	368 (49.1)
	Reply to quiz question	354 (47.2)
	Click on link to the second part of the video	373 (49.7)
17: Group contest on favourite social situation		
	Viewing of others' posts	393 (52.4)[b]
	Posting a picture and message on favourite social situation	91 (12.1)
	Voting on others' posts	144 (19.2)
18: Introduction to substance use resistance skills and quiz on substance use prevalence (alcohol or tobacco) in reference group and normative feedback		
	Click on link to an overview picture	312 (41.6)
	Reply to quiz question	387 (51.6)
19: Quiz on the presence of tobacco advertisements directed to adolescents in everyday life		
	Reply to quiz question	369 (49.2)
	Click on video link	246 (32.8)
20: Quiz on risks of alcohol use		
	Reply to quiz question	386 (51.5)
	Click on website link	238 (31.7)
21: Tailored information on social consequences of alcohol use		
	Click on video link	323 (43.1)
22: Group contest on motivation for abstinence or low-risk alcohol use		
	Viewing of others' posts	393 (52.4)[b]
	Posting a motivational SMS text message	106 (14.1)
	Voting on others' posts	132 (17.6)

[a]The MobileCoach log files do not include a timestamp for this activity. The numbers display the total engagement with this task in the corresponding week 3 and week 6.

[b]The MobileCoach log files do not include a timestamp for this activity. The numbers display the total engagement with this task in the corresponding week.

[c]Information is missing for 1 media object (out of 4) included within the SMS text message replies. This number might underestimate the total engagement of the students with this task.

Measures

After providing informed consent, the students participated in a baseline assessment during a regular class session through which data on potential predictors of engagement and outcome variables were collected. The sociodemographic characteristics that were assessed were gender, age, immigrant background, and type of school (secondary or upper secondary school). We assessed the country of birth of both parents of the students to identify any potential immigrant background. On the basis of this information, the participants were assigned to one of the following categories: (1) neither parent born outside Switzerland, (2) 1 parent born outside Switzerland, or (3) both parents born outside Switzerland.

The health-related variables that were assessed were physical activity, well-being, perceived stress, social skills, problem drinking, tobacco smoking, and cannabis use.

Self-reported moderate to vigorous physical activity was measured using a question derived from the Health Behaviour in School-aged Children study [40]: "Outside school, how many hours a week do you exercise or participate in sports that make you sweat or out of breath?" Well-being was assessed using the World Health Organization–5 Well-being Index [41]. Perceived stress was measured using a single item from the Swiss Juvenir study [42]: "How often have you had the feeling of being overstressed or overwhelmed in the last month?" Participants were asked to indicate their response on a 5-point Likert scale that ranged from *never* to *all the time*. Social skills were assessed by using the brief version of the Interpersonal Competence Questionnaire [43]. Problem drinking was assessed by using the Alcohol Use Disorders Identification Test–Concise (32) with a cut-off of ≥5 based on a large German sample [44]. Tobacco smoking was assessed following the criteria of the Society for Nicotine and Tobacco Research [45] and using the question "Have you smoked a puff within the last 30 days?" Cannabis use was assessed by an item of the Health Behaviour in School-aged Children study [46] addressing the number of cannabis consumption days in the last 30 days.

Use of the program was assessed in terms of the total number of interactions; the number of responses to the weekly SMS text message prompts (quizzes, self-challenges, and individual stress and skills trainings); the number of retrieved media objects within the program (videos, pictures, and website links); and the number of views, posts, and votes within contests. This information was available through the log files of the MobileCoach system.

The subjective experience was assessed at 6-month follow-up by asking the students to report on how attentively they had read the SMS text messages, with the possible answers *thoroughly*, *quickly*, and *not at all*. The answers *quickly* and *not at all* were merged into 1 group.

Statistical Analyses

Descriptive statistics were used to examine participant use of the *SmartCoach* program. Bivariate correlations were performed to evaluate relationships among the use variables. The use variables correlated with each other with scores from r=0.63 (moderate) and r=0.90 (high). For further analysis, the following use outcomes were selected:

1. Total number of interactions with the program, referred to as overall engagement and encompassing all components of the intervention.
2. Total number of quizzes answered.
3. Total number of media objects retrieved.
4. Total number of interactions within contests.

The rationale for selecting the variables 2, 3, and 4 was based on the stipulated effort made by the participant from low (answering a quiz question by typing a predefined letter) to middle (downloading and watching a video) to high (thinking about their own post or voting on other posts within a contest). All use variables were evaluated for normality using Q-Q plots and Shapiro–Wilk tests. According to these checks, all variables demonstrated nonnormal distributions. Finally, associations between overall use and subjective experience were examined.

For detecting relevant individual characteristics associated with engagement (use and subjective experience) with the program, a conditional inference tree, referred to throughout the paper also as decision tree, was estimated [47]. In short, a decision tree follows these steps iteratively:

1. First, it tests for all input variables for which the variable is independent from the response variable (=null hypothesis). If this null hypothesis is not rejected for any input variable, the decision tree stops. Otherwise, it selects the input variable that shows the strongest dependency.
2. On the basis of the selected input variable, it splits the sample into 2 groups.
3. It repeats steps 1 and 2 in both groups until the process stops or until the groups become too small [48].

The advantages of decision trees are as follows: (1) outcomes and residuals do not need to meet assumptions about their distribution, (2) the findings are easy to interpret, and (3) larger amounts of predictors as well as their interactions can be tested simultaneously [47]. The main disadvantages of decision trees are as follows: (1) variable selection is affected by order effects known also from other stepwise variable selection approaches applied within regression analysis, and (2) the strong dependency of the chosen predictors and cut points from the distribution of the observations in the given sample.

Thus, to assess the stability of decision trees, a toolkit of graphical illustrations was used based on resampling [49]. The basic idea of this method is the following: it repeatedly draws new, artificial data sets—so-called bootstrap samples—from the original data set and constructs a decision tree in the new data set. Each bootstrap sample is regarded as a plausible outcome if the study were to be repeated in a new sample. If a decision tree in the bootstrap sample has a similar structure and also selects a similar set of predictor variables, this is interpreted as an indication that these variables have a stable relationship with the outcome variable. In this study, 500 bootstrap samples were drawn for each of the engagement outcomes, and it was investigated whether the predictors that were selected in the original trees were also consistently selected for splitting in the resampled data sets and how often (on average) they were selected for prediction.

In a last step, associations among the engagement outcomes, as mentioned previously, and pre–post changes in the primary and secondary outcomes of the original study were examined. For binary outcomes, follow-up values were included as dependent variables and baseline values, and engagement outcomes (one at a time) were included as independent variables. For continuous outcomes, differences from baseline to 6-month follow-up were included as dependent variables, whereas independent variables were the same as for binary outcomes.

R software (version 3.6.3; The R Foundation for Statistical Computing) and the packages partykit [50,51] and stablelearner [49] were used to perform recursive partitioning and stability assessment, whereas SPSS software (version 22; IBM Corp)

was used for all other analyses. All statistical tests were 2-tailed, with $P<.05$ set as the criterion for statistical significance.

Results

Participants

Table 2 presents baseline characteristics that were used to predict engagement. Of the 750 participants analyzed for this study, 423 (56.4%) were girls. The reported mean age was 15.4 (SD 1.0) years. Almost half of the participants (361/750, 48.2%) reported either a 1-sided or 2-sided immigrant background. Three-fourth (585/750, 78%) of the participants were recruited at an upper secondary school, and the remaining one-fourth (165/750, 22%) were recruited at a secondary school. As expected for this cohort of younger students, only some of the participants reported problem drinking (114/750, 15.2%) and tobacco (91/750, 12.1%) or cannabis use (106/750, 14.1%) 30 days before baseline assessment. The participants reported feeling rather stressed before baseline (mean 2.9, SD 0.9; Q1, Q3=2, 4). In addition, the mean Well-being Index score was 52.9 (SD 17.3; Q1, Q3=40, 76) out of 100 possible points. On the basis of the Interpersonal Competence Questionnaire, brief form, the mean level of interpersonal competences was rather high.

Table 2. Baseline characteristics of the study sample (N=750).

Variable	Values
Sex, n (%)	
Male	327 (43.6)
Female	423 (56.4)
Age (years), mean (SD)	15.4 (1.0)
Number of physically active hours per week[a], mean (SD)	4.1 (3.5)
Immigration background, n (%)	
No immigration background	389 (51.9)
One parent born outside Switzerland	173 (23.1)
Both parents born outside Switzerland	188 (25.1)
Type of school, n (%)	
Secondary school	165 (22)
Upper secondary school	585 (78)
Tobacco smoking before baseline, yes, n (%)	91 (12.1)
Problem drinking before baseline, yes, n (%)	114 (15.2)
Cannabis use before baseline, yes, n (%)	106 (14.1)
Perceived stress on a scale of 1-5, mean (SD)	2.9 (0.9)
Well-being (WHO[b]-5 Well-being Index) on a scale of 1-100, mean (SD)	52.9 (17.3)
Interpersonal competences (ICQ-10[c]) on a scale of 5-20, mean (SD)	14.9 (2.2)

[a]Missing values: n=1.

[b]WHO: World Health Organization.

[c]ICQ-10: Interpersonal Competence Questionnaire, brief form.

Use of Different Program Components

Table 3 summarizes different program use characteristics across the sample. Of the 750 students, 40 (5.3%) discontinued the intervention actively by sending an SMS text message to the program. Of these 40 students, 17 (43%) signed off at program start, whereas the rest (23/40, 58%) signed off somewhere between the first and 16th program weeks.

The participants replied to an average of 23.6 (SD 15.9) of 50 program prompts. They interacted with the program mainly through SMS text message and the least by participating in a picture or message contest involving all intervention participants. They answered a mean of 12 (SD 7.7) of 21 SMS text message prompts. On average, the participants downloaded 8.8 (SD 6.9) of 20 media objects received. In all, 3 contests were prompted throughout the program, each of them including 3 steps (viewing, posting, and voting). Although 50% (375/750) of the participants viewed at least two contests, only a few participants accepted the invitation to post pictures or messages or voted on them within the contests. No participant posted something during all 3 contests.

Table 1 displays the number of students engaging with the prompted activities of the program each week. In general, engagement with the program decreased over time, as can be observed for the most popular activity, replying to quiz questions (week 1=525/750, 70%; week 20=386/750, 51.5%). However, the trend is not linear, and there are weeks with engagement peaks depending on the topic (eg, in week 8, 372/750, 49.6%,

clicked on the link to an overview of the social skills topic, whereas only 194/750, 25.9%, did so in week 1; in week 9, 716/750, 95.5%, replied to the SMS text messages with options for improving their social skills).

Table 3. Use of program components.

Engagement variables	Values, mean (SD)	Values, median (Q1[a], Q3[b])	Range
Total number of interactions	23.6 (15.9)	26.5 (8, 38)	0-50
SMS text messages			
Quizzes	6.2 (4.0)	7 (2, 7)	0-11
Stress-trainings	1.3 (0.9)	2 (0, 2)	0-2
Self-challenges	3.4 (2.4)	4 (1, 6)	0-6
Skill-trainings	1.1 (0.9)	1 (0, 2)	0-2
Total use of text messages	12.0 (7.7)	14 (4, 19)	0-21
Media objects			
Videos	5.1 (4.0)	5 (1, 9)	0-12
Pictures	1.9 (1.5)	2 (1, 3)	0-4
Website links	1.7 (1.7)	1 (0, 3)	0-6
Total use of media objects	8.8 (6.9)	9 (2, 15)	0-20
Picture and SMS text message contests			
Views	1.8 (1.3)	2 (0, 3)	0-3
Posts	0.3 (0.4)	0 (0, 1)	0-1
Votes	0.7 (0.9)	0 (0, 1)	0-3
Total use of contests	2.9 (2.5)	3 (0, 5)	0-9

[a]Q1: quartile 1.

[b]Q3: quartile 3.

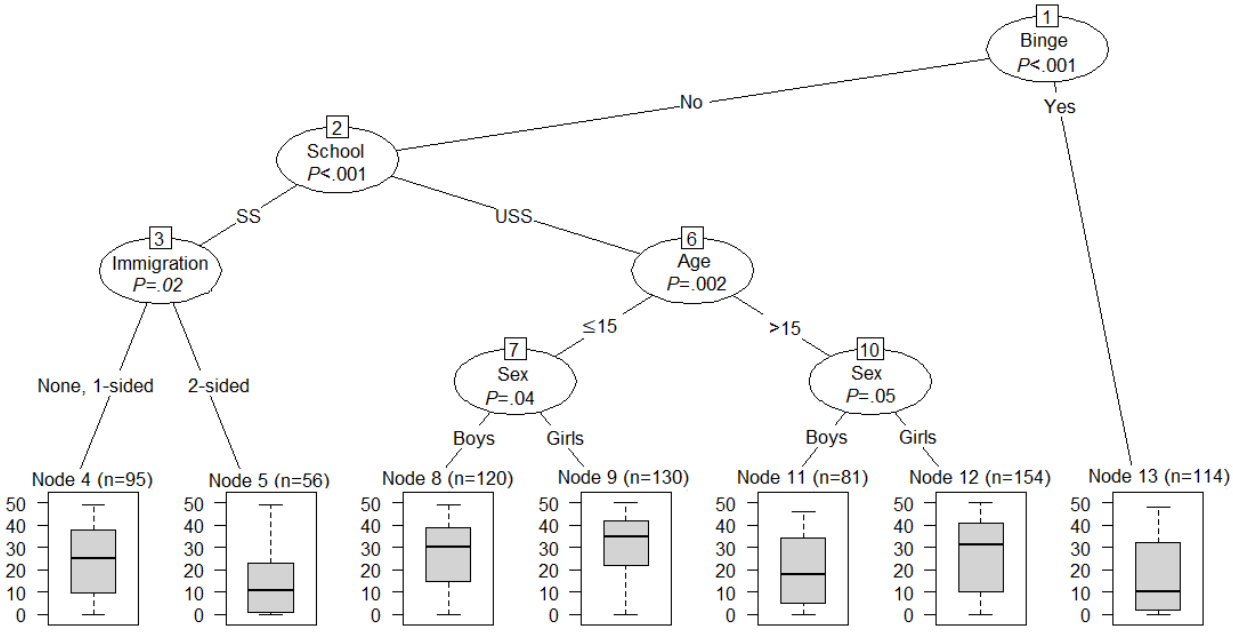
Predictors of Engagement

Predictors of Total Use

Figure 3 depicts a decision tree for the prediction of overall use of the *SmartCoach* program. Baseline characteristics used for prediction of overall use were problem drinking before baseline, school type, immigration background, age, and sex (Figure 3). The decision tree can be read as follows: students who reported problem drinking before baseline (node 13) and students who did not drink in a problematic manner before baseline but attended a secondary school and reported a 2-sided immigration background (node 5) were expected to show low use of the program. The highest use was expected for girls aged ≤15 years who attended an upper secondary school and who reported no alcohol use or moderate alcohol use before baseline (node 9). Girls at upper secondary schools used the program more than boys at upper secondary schools. At secondary schools, use of the program was similar across both sexes.

Multimedia Appendix 1 gives an insight into the stability of this decision tree. Each column indicates how often each predictor variable was selected in the 500 bootstrap samples. Whereas variables selected in the original tree are presented as black columns, variables that were not selected in the original tree are presented as white columns. The predictors problem drinking and school type were selected very often (approximately 80% of the time) for the prediction of overall use and therefore showed a relatively stable relationship with the outcome. However, the predictors sex, age, and immigration background were only represented in less than half of the trees generated based on the bootstrap samples, which makes them less reliable for prediction compared with the predictors problem drinking and school type. Tobacco smoking, cannabis use, perceived stress, and interpersonal competences were not represented in the original tree but appeared in 10%-15% of the other decision trees. This indicates that although those variables seem to carry some information that could be useful for predicting overall use, they are not predominant.

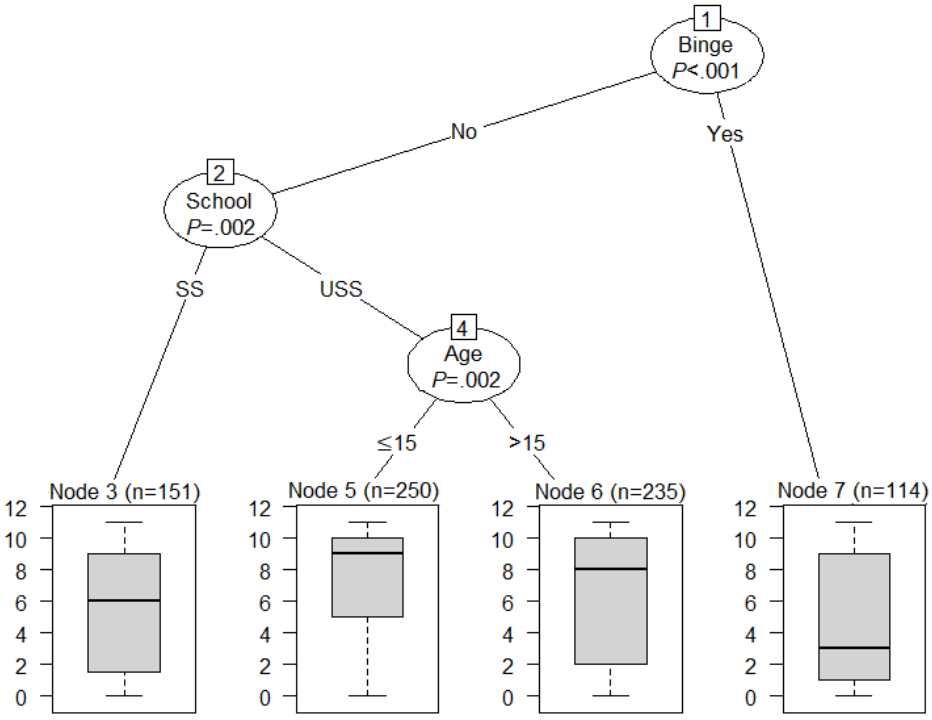
Figure 3. Decision tree with predictors of overall use of the *SmartCoach* program. Binge: problem drinking; Immigration: immigration background; School: school type; SS: secondary school; USS: upper secondary school.



Predictors of Quiz Use

Figure 4 depicts a decision tree for quiz use, encompassing the following predictors: problem drinking, school type, and age. However, the stability check (Multimedia Appendix 2) implies that only problem drinking showed a stable relationship with the outcome, being chosen in 80% (400/500) of the subsamples for prediction. School type was chosen only in approximately 50% of the 500 decision trees and age in less than 50% and almost as often as immigration background, which was not depicted in the original tree. To sum up, this decision tree should be read as follows: nondrinkers at baseline showed higher use of quizzes, whereas students who reported problem drinking at baseline showed lower use.

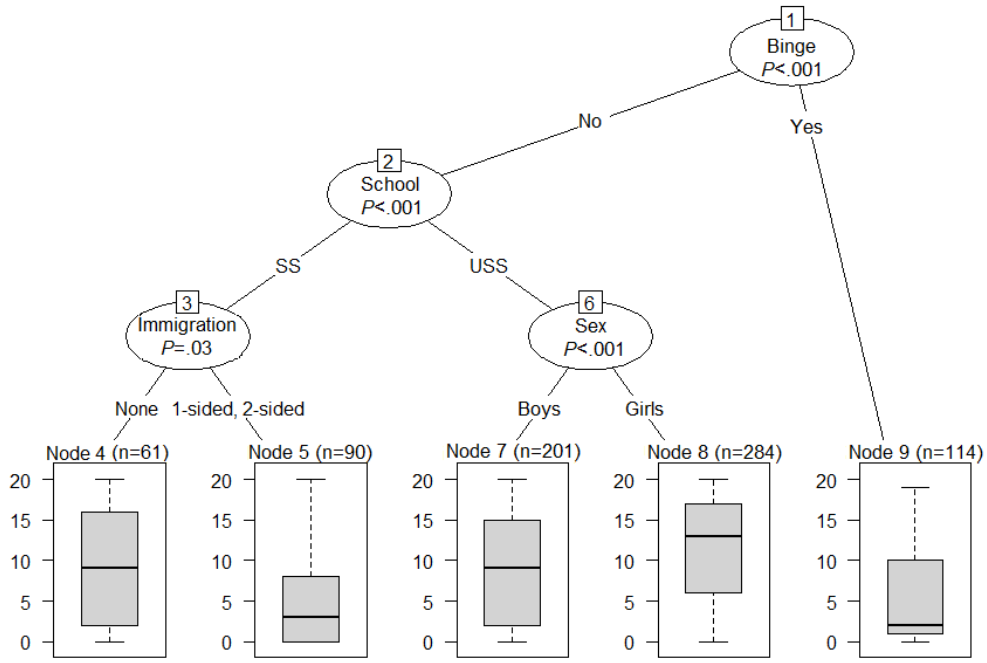
Figure 4. Decision tree for the prediction of use of quizzes. Binge: problem drinking; School: school type; SS: secondary school; USS: upper secondary school.



Predictors of Media Use

Figure 5 illustrates a decision tree for use of media objects, encompassing the following predictors: problem drinking, school type, sex, and immigration background. However, the stability check (Multimedia Appendix 3) implies that immigration background has a less stable relationship with the use of media objects because it was only chosen in approximately 30% of the decision trees and almost as often as tobacco smoking, age, and cannabis use. The other predictor variables showed a more stable relationship. This once again could be read as follows: students who reported problem drinking at baseline were expected to use media objects the least, whereas regular use was expected across nondrinkers assisting different school types with the only difference being that girls at upper secondary schools used this program component more than the rest.

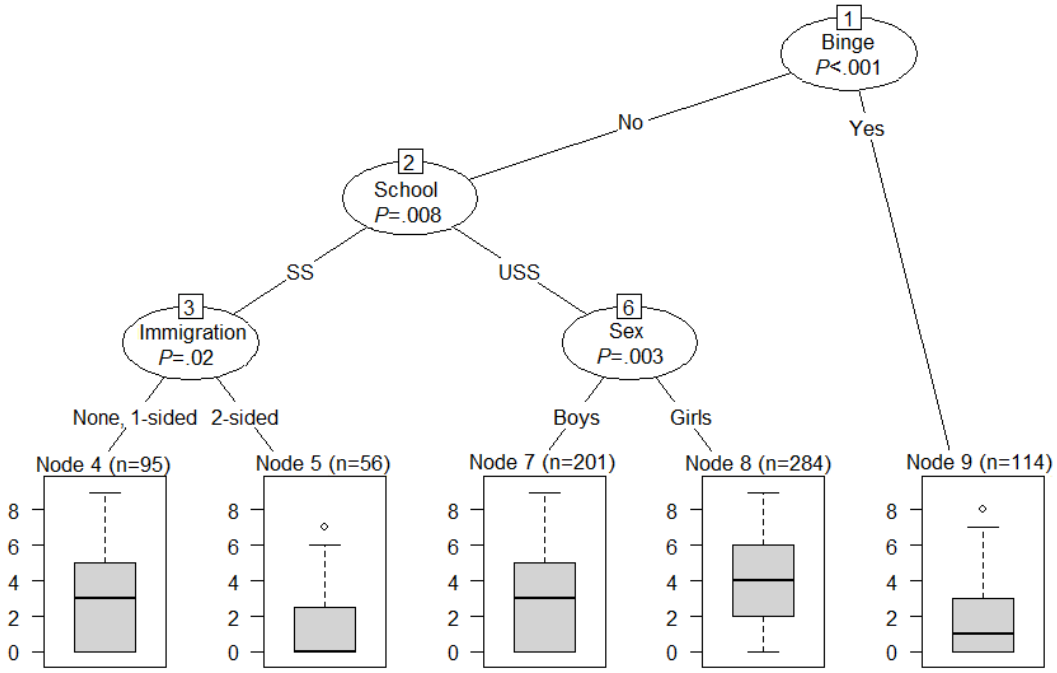
Figure 5. Decision tree for the prediction of use of media objects. Binge: problem drinking; Immigration: immigration background; School: school type; SS: secondary school; USS: upper secondary school.



Predictors of Contest Use

Figure 6 depicts a decision tree for use of contests, encompassing the following predictors: problem drinking, school type, sex, and immigration background. The tree looks very similar to the one for media objects, with the only difference being a split in the immigration background. However, the stability check (Multimedia Appendix 4) implies that none of these predictors showed a highly stable relationship with the use of contests because they only appear approximately 50% of the time or even less. This decision tree should therefore not be considered for further interpretation.

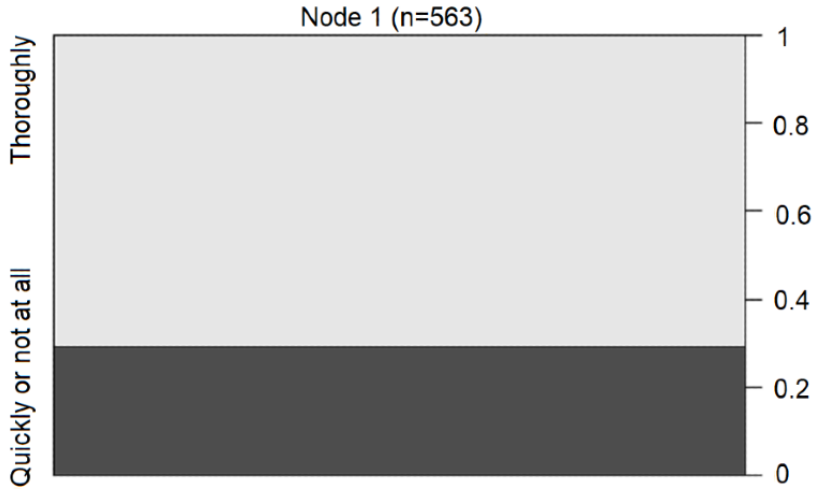
Figure 6. Decision tree for the prediction of use of contests. Binge: problem drinking; Immigration: immigration background; School: school type; SS: secondary school; USS: upper secondary school.



Predictors of Subjective Experience

Figure 7 depicts the result of the decision tree for predicting the subjective experience of students with the program. As it becomes clear from Figure 7 and from Multimedia Appendix 5, none of the included predictors showed a stable relationship with the outcome.

Figure 7. Decision tree for the prediction of subjective experience, which has no splits and consists of a single node.



Associations Among Use, Subjective Experience, and Changes in Outcomes

Overall use of the program was not predictive in our complete case analysis for problem drinking (odds ratio [OR] 0.99, 95% CI 0.03-0.09; *P*=.65) or cannabis smoking (OR 0.98, 95% CI 0.97-1.01; *P*=.20) at follow-up or for changes in interpersonal competences (β*s*=.2; t_{571}=0.41; *P*=.68).

A tendency was observed for the outcome tobacco use, where participants who interacted more frequently with the program showed lower odds of having smoked at follow-up (OR 0.98, 95% CI 0.96-1.00; *P*=.05; R^2=0.18) compared with baseline. Ancillary analysis on tobacco smoking revealed that use of contests compared with the rest of the program components lowered the odds of having smoked at follow-up (OR 0.86, 95% CI 0.76-0.98; *P*=.02; R^2=0.18). Finally, a significant quadratic association was established between overall use (β*s*=.39; t_{586}=2.66; *P*=.008; R^2=0.24) of the program and pre–post changes in well-being assessed using the World Health Organization–5 Well-being Index. All participants showed an

increase in well-being from baseline to 6-month follow-up, but the mean increase was greater for lower- and higher-engaged participants. The highest increase in well-being was observed for participants who participated in almost all program prompts.

Subjective experience (assessed by asking the students to report on how attentively they had read the SMS text messages) was predictive for problem drinking at follow-up. Students who read the SMS text messages more attentively were significantly less at risk for problem drinking at follow-up (OR 0.43, 95% CI 1.29-3.41; P=.003; R^2=0.19).

Discussion

Principal Findings

Using a proactively recruited sample of young adolescents in secondary and upper secondary schools, this study examined (1) the use of, and subjective experience with, a digital life skills intervention; (2) the associations between program engagement and adolescent characteristics; (3) the stability of these associations using a resampling method; and (4) the associations between program engagement and changes in health-related outcomes.

The main findings are as follows:

1. Adolescents took part in half of the interactions prompted by the program, with SMS text messages being the most used and contests being the least used components.
2. Of all adolescent characteristics included in the decision tree analysis, the following were chosen most often as a predictor of engagement: problem drinking, school type, age, sex, and immigration background. In almost all decision trees, adolescents who reported problematic patterns of alcohol use at baseline were expected to use the program the least, followed by adolescents attending a secondary school and reporting a 2-sided immigration background. Often, a higher use of the program and its different components was expected for girls and younger adolescents (aged ≤15 years).
3. Predictors included in the original trees were often considered unstable for prediction after resampling. The most stable predictor for overall and component use after resampling was problem drinking before baseline (relative frequency [RF] of approximately 80% in 3 out of 4 trees), followed by school type (RF≥80% in 2 out of 4 trees). Predictors of contest use and subjective experience were especially unstable (RF≤50%).
4. Adolescents who used the contests more intensively were more likely to be nonsmokers at follow-up compared with those who did not. In addition, participants who interacted the most and the least with the program were more likely to increase their well-being from baseline to 6-month follow-up. Finally, adolescents who read the SMS text messages more attentively were less likely to drink in a problematic manner at follow-up.

This is the first study to examine engagement with a mobile phone–based life skills intervention among younger adolescents. Similar results were found for 2 other mobile phone–based interventions directed at older adolescents, namely that SMS text message prompts and especially quizzes were the most used, whereas contests, where a message has to be produced by the participant, were the least used [12,29]. However, this study showed that contest use was associated with the probability of being a nonsmoker at follow-up, and a recent review on digital mental health interventions [21] concluded that program components enabling interaction with peers were the most engaging, which is why this component should not be eliminated from the intervention without further examination.

The results of this study coincide with previous research, identifying younger adolescents [27,29], women [21], and adolescents without an immigration background [29] to engage more with digital interventions aimed at promoting mental health or preventing substance use. However, this study is the first to not only examine predictors of engagement, but to also analyze the stability of these predictors, and in doing so, to overcome order effects in variable selection and dependency from the recruited sample known from standard regression analysis. As a result, sex, age, and immigration background were detected as rather unstable predictors, which is why associations between these adolescent characteristics and engagement outcomes should be considered only with caution.

The most stable predictor over all decision trees was problem drinking reported at baseline. Generally speaking, adolescents who already showed patterns of problematic alcohol use at baseline were expected to engage the least with the program. The same association could not be found for tobacco or cannabis use, probably because the included items measured prevalence instead of severity of use, but this remains a question for future research. In other words, for those adolescents who are already involved in risky substance use, an unspecific intervention might not be appealing because they cannot infer their benefit. This would coincide with previous research, which found a tobacco-specific intervention to be more engaging for adolescents with higher self-perceived benefit in quitting smoking [29] and a mental health intervention to be more engaging for adolescents with higher levels of depressive symptoms [27]. Future qualitative work is needed to ascertain if digital life skills interventions should be improved for adolescents who are already involved in risky substance use or if a problem-specific digital intervention could be of more interest for this subgroup of adolescents.

Another rather stable predictor was school type, which can be interpreted as a proxy for educational level in this study. The results indicate that the mobile phone–based life skills intervention was more engaging for upper secondary school students than for secondary school students. This suggests that the intervention could have been still too demanding for students with lower educational attainment. Specifically, media objects—but not quizzes—could have been too demanding for those adolescents. However, interaction with quizzes alone was not found to be associated with changes in health-related outcomes, indicating that despite being engaging for users, quizzes are less useful for fostering behavior change and more useful for ensuring program use and for preventing program dropout. New ways to make media objects more engaging, including for those with lower educational attainment, must be explored. A novel chatbot-based life skills intervention that was

developed in a participatory manner with 20 adolescents [16] showed that simple cartoon videos were more appealing than text information and more useful for fostering self-reflection on individual life skills if they were presented at the beginning of a session. Cooperating with organizations that already produce engaging and relevant digital content for adolescents, as planned by the developers of a novel mobile-based intervention to support social emotional learning and identity development in adolescents from low-resource contexts [17], could also increase the use of such components within digital life skills interventions.

The results of this study suggest that there is a positive association between engagement and intended outcomes. A higher use of contests and the overall program were associated with intended tobacco use and well-being outcomes. In addition, reading the text messages more attentively was associated with intended alcohol use at follow-up. However, low use of the program also predicted higher well-being at follow-up compared with baseline. These mixed results could stem from the different reach at follow-up of more-engaged compared with less-engaged participants, which is a rather well-known problem from previous studies [29,52] or from the fact that less-engaged participants received sufficient support to facilitate intrinsic motivation to boost well-being [53]. Both explanations are plausible when looking into previous studies with adults, where less-engaged participants with access to mobile phone–based smoking reduction programs reported greater changes at follow-up [54,55]. Challenges for future research remain finding ways to also assess lower-engaged participants at follow-up. The strategy of this study—reimbursement of CHF 10 (US $10.90) for participation at each follow-up and up to 5 follow-up calls—seems to be insufficient. A possible way to combat this problem for studies recruiting in school settings could be to conduct follow-up assessments within regular school sessions.

Limitations

The findings of this study must be interpreted in view of its limitations. First, answers to weekly prompts were rewarded with credits, and credits were linked to a prize draw. Second, as already emphasized in other studies [29,55], the quantity and quality of answers to prompts could differ (eg, an adolescent who answers all quiz questions but answers them all wrong). Rather than just analyzing registered events, future qualitative work should investigate whether the content of answers or contest posts is associated with treatment outcomes. Furthermore, the results of this study rely on a convenience sample, and the findings might not be generalizable to the entire population. Finally, this study relied on self-report data, which bears the risk that the results may have been influenced by social desirability.

Conclusions

In summary, in our study, adolescents who did not drink in a problematic manner before program start frequently engaged with a mobile phone–based life skills intervention, regardless of their sex, age, and immigration background. Further efforts should be undertaken to reach adolescents through digital life skills interventions before they become involved with risky substance use. Digitally delivered life skills interventions must carefully consider how the proportion between engaging and change-modeling components should be weighed to comply equally with equity standards and intended intervention outcomes. In addition, future studies could go a step further and make assumptions about *effective* engagement but only in combination with strategies that are able to reduce attrition bias at follow-up.

Acknowledgments

Funding for this project was provided by the Swiss National Science Foundation (grant number 10001C_179222/1).

Conflicts of Interest

RPC and SH were also involved in the development of the intervention program *SmartCoach*. RJ and TK are affiliated with the Centre for Digital Health Interventions, a joint initiative of the Department of Management, Technology, and Economics at ETH Zurich and the Institute of Technology Management at the University of St Gallen, which is funded in part by the Swiss health insurer, CSS. TK is also cofounder of Pathmate Technologies, a university spin-off company that creates and delivers digital clinical pathways. However, neither CSS nor Pathmate Technologies was involved in the design or results analysis of the presented study.The funding institution did not influence the design and conduct of the study; the management, analysis, or interpretation of data; or the preparation, review, or approval of the manuscript.

Multimedia Appendix 1

Corresponding assessment for predictor stability of the decision tree for the prediction of overall use. Binge: problem drinking; Cannabis: cannabis use; ICQ: Interpersonal Competence Questionnaire score; Immigration: immigration background; Physical: physical activity; School: school type; Stress: perceived stress; Tobacco: tobacco use; WHO: World Health Organization–5 Well-being Index score.

[PNG File , 24 KB - jmir_v24i1e28638_app1.png]

Multimedia Appendix 2

Corresponding assessment for predictor stability of the decision tree for the prediction of use of quizzes. Binge: problem drinking; Cannabis: cannabis use; ICQ: Interpersonal Competence Questionnaire score; Immigration: immigration background; Physical: physical activity; School: school type; Stress: perceived stress; Tobacco: tobacco use; WHO: World Health Organization–5 Well-being Index score.
[PNG File , 23 KB - jmir_v24i1e28638_app2.png]

Multimedia Appendix 3
Corresponding assessment for predictor stability of the decision tree for the prediction of use of media objects. Binge: problem drinking; Cannabis: cannabis use; ICQ: Interpersonal Competence Questionnaire score; Immigration: immigration background; Physical: physical activity; School: school type; Stress: perceived stress; Tobacco: tobacco use; WHO: World Health Organization–5 Well-being Index score.
[PNG File , 23 KB - jmir_v24i1e28638_app3.png]

Multimedia Appendix 4
Corresponding assessment for predictor stability of the decision tree for the prediction of use of contests. Binge: problem drinking; Cannabis: cannabis use; ICQ: Interpersonal Competence Questionnaire score; Immigration: immigration background; Physical: physical activity; School: school type; Stress: perceived stress; Tobacco: tobacco use; WHO: World Health Organization–5 Well-being Index score.
[PNG File , 23 KB - jmir_v24i1e28638_app4.png]

Multimedia Appendix 5
Corresponding assessment for predictor stability of the decision tree for the prediction of subjective experience. Binge: problem drinking; Cannabis: cannabis use; ICQ: Interpersonal Competence Questionnaire score; Immigration: immigration background; Physical: physical activity; School: school type; Stress: perceived stress; Tobacco: tobacco use; WHO: World Health Organization–5 Well-being Index score.
[PNG File , 23 KB - jmir_v24i1e28638_app5.png]

References

1. Costello E, Copeland W, Angold A. Trends in psychopathology across the adolescent years: what changes when children become adolescents, and when adolescents become adults? J Child Psychol Psychiatry 2011 Oct;52(10):1015-1025 [FREE Full text] [doi: 10.1111/j.1469-7610.2011.02446.x] [Medline: 21815892]
2. Inchley J, Currie D, Budisavljevic S, Torsheim T, Jåstad A, Cosma A, et al. Spotlight on Adolescent Health and Well-being. Findings From the 2017/2018 Health Behaviour in School-Aged Children (HBSC) Survey in Europe and Canada. International Report. Volume 1. Key Findings. Geneva: World Health Organization; 2020.
3. Jordan MD, Schneider E, Eichenberger Y, Kretschmann A. La consommation de substances psychoactives des 11 à 15 ans en Suisse - Situation en 2018 et évolutions depuis 1986 - Résultats de l?étude Health Behaviour in School-aged Children (HBSC). Lausanne, Switzerland. 2019. URL: https://www.addictionsuisse.ch/fileadmin/user_upload/DocUpload/HBSC-2018-rr-100.pdf [accessed 2021-12-16]
4. Stockings E, Hall WD, Lynskey M, Morley KI, Reavley N, Strang J, et al. Prevention, early intervention, harm reduction, and treatment of substance use in young people. Lancet Psychiatry 2016 Mar;3(3):280-296. [doi: 10.1016/s2215-0366(16)00002-x]
5. Thomas RE, McLellan J, Perera R. School-based programmes for preventing smoking. Evid Based Child Health 2013 Sep 18;8(5):1616-2040. [doi: 10.1002/ebch.1937]
6. Foxcroft D, Tsertsvadze A. Universal school-based prevention programs for alcohol misuse in young people. Cochrane Database Syst Rev 2011 May 11(5):CD009113. [doi: 10.1002/14651858.CD009113] [Medline: 21563171]
7. Weichold K, Blumenthal A. Long-term effects of the life skills program IPSY on substance use: results of a 4.5-year longitudinal study. Prev Sci 2016 Jan;17(1):13-23. [doi: 10.1007/s11121-015-0576-5] [Medline: 26202801]
8. Velasco V, Griffin KW, Botvin GJ, Corrado Celata and Gruppo LST Lombardia. Preventing adolescent substance use through an evidence-based program: effects of the Italian adaptation of life skills training. Prev Sci 2017 May 28;18(4):394-405. [doi: 10.1007/s11121-017-0776-2] [Medline: 28353126]
9. Domitrovich C, Durlak J, Staley K, Weissberg R. Social-emotional competence: an essential factor for promoting positive adjustment and reducing risk in school children. Child Dev 2017 Mar;88(2):408-416. [doi: 10.1111/cdev.12739] [Medline: 28213889]
10. Gates S, McCambridge J, Smith L, Foxcroft D. Interventions for prevention of drug use by young people delivered in non-school settings. Cochrane Database Syst Rev 2006 Jan 25(1):CD005030. [doi: 10.1002/14651858.CD005030.pub2] [Medline: 16437511]
11. Vroom EB, Massey OT, Yampolskaya S, Levin BL. The impact of implementation fidelity on student outcomes in the life skills training program. School Ment Health 2019 Jun 20;12(1):113-123. [doi: 10.1007/s12310-019-09333-1]

12. Haug S, Paz Castro R, Meyer C, Filler A, Kowatsch T, Schaub M. A mobile phone-based life skills training program for substance use prevention among adolescents: pre-post study on the acceptance and potential effectiveness of the program, Ready4life. JMIR Mhealth Uhealth 2017 Oct 04;5(10):e143 [FREE Full text] [doi: 10.2196/mhealth.8474] [Medline: 28978498]
13. Haug S, Paz Castro R, Wenger A, Schaub M. Efficacy of a smartphone-based coaching program for addiction prevention among apprentices: study protocol of a cluster-randomised controlled trial. BMC Public Health 2020 Dec 14;20(1):1910 [FREE Full text] [doi: 10.1186/s12889-020-09995-6] [Medline: 33317488]
14. Haug S, Paz Castro R, Wenger A, Schaub MP. Efficacy of a mobile phone-based life-skills training program for substance use prevention among adolescents: study protocol of a cluster-randomised controlled trial. BMC Public Health 2018 Sep 10;18(1):1102 [FREE Full text] [doi: 10.1186/s12889-018-5969-5] [Medline: 30200928]
15. Field M, Campbell F, Hock E, Wong R. Rapid Evidence Review Digital interventions to reduce alcohol-related harm: a rapid horizon scanning review. Alcohol Change UK. 2019. URL: https://s3.eu-west-2.amazonaws.com/files.alcoholchange.org.uk/documents/Field-Campbell-Hock-Wong-Digital-Interventions-FINAL-REPORT.pdf [accessed 2021-12-16]
16. Gabrielli S, Rizzi S, Carbone S, Donisi V. A chatbot-based coaching intervention for adolescents to promote life skills: pilot study. JMIR Hum Factors 2020 Feb 14;7(1):e16762 [FREE Full text] [doi: 10.2196/16762] [Medline: 32130128]
17. Cherewick M, Lebu S, Su C, Richards L, Njau PF, Dahl RE. Study protocol of a distance learning intervention to support social emotional learning and identity development for adolescents using interactive mobile technology. Front Public Health 2021 Jan 28;9:623283 [FREE Full text] [doi: 10.3389/fpubh.2021.623283] [Medline: 33585394]
18. Haug S, Paz Castro R, Wenger A, Schaub MP. A mobile phone-based life-skills training program for substance use prevention among adolescents: cluster-randomized controlled trial. JMIR Mhealth Uhealth 2021 Jul 13;9(7):e26951 [FREE Full text] [doi: 10.2196/26951] [Medline: 34255703]
19. Ellickson PL. You've shown the program model is effective. Now what? New Dir Youth Dev 2014;2014(141):95-105, 13. [doi: 10.1002/yd.20089] [Medline: 24753281]
20. MacDougall S, Jerrott S, Clark S, Campbell L, Murphy A, Wozney L. Text message interventions in adolescent mental health and addiction services: scoping review. JMIR Ment Health 2021 Jan 08;8(1):e16508 [FREE Full text] [doi: 10.2196/16508] [Medline: 33416504]
21. Garrido S, Millington C, Cheers D, Boydell K, Schubert E, Meade T, et al. What works and what doesn't work? A systematic review of digital mental health interventions for depression and anxiety in young people. Front Psychiatry 2019 Nov 13;10:759 [FREE Full text] [doi: 10.3389/fpsyt.2019.00759] [Medline: 31798468]
22. Struthers A, Charette C, Bapuji S, Winters S, Ye X, Metge C, et al. The acceptability of e-mental health services for children, adolescents, and young adults: a systematic search and review. Can J Community Ment Health 2015 Jul 01;34(2):1-21. [doi: 10.7870/cjcmh-2015-006]
23. Kazemi DM, Borsari B, Levine MJ, Li S, Lamberson KA, Matta LA. A systematic review of the mhealth interventions to prevent alcohol and substance abuse. J Health Commun 2017 May 10;22(5):413-432 [FREE Full text] [doi: 10.1080/10810730.2017.1303556] [Medline: 28394729]
24. Donkin L, Christensen H, Naismith SL, Neal B, Hickie IB, Glozier N. A systematic review of the impact of adherence on the effectiveness of e-therapies. J Med Internet Res 2011 Aug 05;13(3):e52 [FREE Full text] [doi: 10.2196/jmir.1772] [Medline: 21821503]
25. Perski O, Blandford A, West R, Michie S. Conceptualising engagement with digital behaviour change interventions: a systematic review using principles from critical interpretive synthesis. Transl Behav Med 2017 Jun;7(2):254-267 [FREE Full text] [doi: 10.1007/s13142-016-0453-1] [Medline: 27966189]
26. Achilles M, Anderson M, Li S, Subotic-Kerry M, Parker B, O'Dea B. Adherence to e-mental health among youth: Considerations for intervention development and research design. Digit Health 2020;6:2055207620926064 [FREE Full text] [doi: 10.1177/2055207620926064] [Medline: 32547775]
27. Calear AL, Christensen H, Mackinnon A, Griffiths KM. Adherence to the MoodGYM program: outcomes and predictors for an adolescent school-based population. J Affect Disord 2013 May;147(1-3):338-344. [doi: 10.1016/j.jad.2012.11.036] [Medline: 23245469]
28. Rose T, Barker M, Maria JC, Morrison L, Lawrence W, Strömmer S, et al. A systematic review of digital interventions for improving the diet and physical activity behaviors of adolescents. J Adolesc Health 2017 Dec;61(6):669-677 [FREE Full text] [doi: 10.1016/j.jadohealth.2017.05.024] [Medline: 28822682]
29. Paz Castro R, Haug S, Filler A, Kowatsch T, Schaub MP. Engagement within a mobile phone-based smoking cessation intervention for adolescents and its association with participant characteristics and outcomes. J Med Internet Res 2017 Nov 01;19(11):e356 [FREE Full text] [doi: 10.2196/jmir.7928] [Medline: 29092811]
30. Whitton A, Proudfoot J, Clarke J, Birch M, Parker G, Manicavasagar V, et al. Breaking open the black box: isolating the most potent features of a web and mobile phone-based intervention for depression, anxiety, and stress. JMIR Ment Health 2015;2(1):e3 [FREE Full text] [doi: 10.2196/mental.3573] [Medline: 26543909]

31. Clarke G, Kelleher C, Hornbrook M, Debar L, Dickerson J, Gullion C. Randomized effectiveness trial of an Internet, pure self-help, cognitive behavioral intervention for depressive symptoms in young adults. Cogn Behav Ther 2009 May 13;38(4):222-234 [FREE Full text] [doi: 10.1080/16506070802675353] [Medline: 19440896]
32. Mohr DC, Burns MN, Schueller SM, Clarke G, Klinkman M. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. Gen Hosp Psychiatry 2013 Aug;35(4):332-338 [FREE Full text] [doi: 10.1016/j.genhosppsych.2013.03.008] [Medline: 23664503]
33. Christensen H, Mackinnon A. The law of attrition revisited. J Med Internet Res 2006 Sep 29;8(3):20-21 [FREE Full text] [doi: 10.2196/jmir.8.3.e20] [Medline: 17032636]
34. Yardley L, Spring BJ, Riper H, Morrison LG, Crane DH, Curtis K, et al. Understanding and promoting effective engagement with digital behavior change interventions. Am J Prev Med 2016 Nov;51(5):833-842. [doi: 10.1016/j.amepre.2016.06.015] [Medline: 27745683]
35. Solar C, Halat A, MacLean R, Rajeevan H, Williams D, Krein S, et al. Predictors of engagement in an internet-based cognitive behavioral therapy program for veterans with chronic low back pain. Transl Behav Med 2021 Jun 17;11(6):1274-1282. [doi: 10.1093/tbm/ibaa098] [Medline: 33098304]
36. Ellickson P, McCaffrey D, Ghosh-Dastidar B, Longshore D. New inroads in preventing adolescent drug use: results from a large-scale trial of project ALERT in middle schools. Am J Public Health 2003 Nov;93(11):1830-1836. [doi: 10.2105/ajph.93.11.1830] [Medline: 14600049]
37. Botvin G, Griffin C. Life skills training: preventing substance misuse by enhancing individual and social competence. New Dir Youth Dev 2014;2014(141):57-65. [doi: 10.1002/yd.20086] [Medline: 24753278]
38. Filler A, Kowatsch T, Haug S, Wahle F, Staake T, Fleisch E. MobileCoach: a novel open source platform for the design of evidence-based, scalable and low-cost behavioral health interventions: Overview and preliminary evaluation in the public health context. In: Proceedings of the Wireless Telecommunications Symposium (WTS). 2015 Presented at: Wireless Telecommunications Symposium (WTS); April 15-17, 2015; New York, NY, USA. [doi: 10.1109/wts.2015.7117255]
39. Haug S, Kowatsch T, Paz Castro R, Filler A, Schaub M. Efficacy of a web- and text messaging-based intervention to reduce problem drinking in young people: study protocol of a cluster-randomised controlled trial. BMC Public Health 2014 Aug 07;14:809 [FREE Full text] [doi: 10.1186/1471-2458-14-809] [Medline: 25099872]
40. Inchley J, Currie D, Young T, Samdal O, Torsheim T, Augustson L, et al. Growing up unequal: gender and socioeconomic differences in young people's health and well-being: Health Behaviour in School-aged Children (HBSC) study: international report from the 2013/2014 survey. (Health Policy for Children and Adolescents; No. 7). WHO Regional Office for Europe. 2016. URL: https://tinyurl.com/2f95hnxf [accessed 2021-12-16]
41. Topp C, Østergaard SD, Søndergaard SP, Bech P. The WHO-5 Well-Being Index: a systematic review of the literature. Psychother Psychosom 2015;84(3):167-176 [FREE Full text] [doi: 10.1159/000376585] [Medline: 25831962]
42. Zuviel Stress – zuviel Druck! : Wie Schweizer Jugendliche mit Stress und Leistungsdruck umgehen. Jacobs Foundation.Juvenir-Studie 4.0. 2015. URL: https://jacobsfoundation.org/app/uploads/2017/07/Juvenir-4.0_Kurzfassung_DE_final.pdf [accessed 2021-12-16]
43. Coroiu A, Meyer A, Gomez-Garibello C, Brähler E, Hessel A, Körner A. Brief Form of the Interpersonal Competence Questionnaire (ICQ-15). Eur J Psychol Assess 2015 Oct;31(4):272-279. [doi: 10.1027/1015-5759/a000234]
44. Rumpf H, Wohlert T, Freyer-Adam J, Grothues J, Bischof G. Screening questionnaires for problem drinking in adolescents: performance of AUDIT, AUDIT-C, CRAFFT and POSIT. Eur Addict Res 2013;19(3):121-127. [doi: 10.1159/000342331] [Medline: 23183686]
45. Hughes JR, Keely JP, Niaura RS, Ossip-Klein DJ, Richmond RL, Swan GE. Measures of abstinence in clinical trials: issues and recommendations. Nicot Tob Res 2003 Jan 01;5(1):13-25. [doi: 10.1093/ntr/5.1.13]
46. Marmet S, Archimi A, Windlin B, Jordan MD. Substanzkonsum bei Schülerinnen und Schülern in der Schweiz im Jahr 2014 und Trend seit 1986 : Resultate der Studie “Health Behaviour in School-aged Children” (HBSC). Lausanne, Switzerland. 2015. URL: https://www.hbsc.ch/pdf/hbsc_bibliographie_292.pdf [accessed 2021-12-16]
47. Strobl C, Malley J, Tutz G. An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests. Psychol Methods 2009 Dec;14(4):323-348 [FREE Full text] [doi: 10.1037/a0016973] [Medline: 19968396]
48. Hothorn T, Hornik K, Wien W, Zeileis A. party: a laboratory for recursive partytioning. Cran.R-Project. URL: https://cran.r-project.org/web/packages/party/vignettes/party.pdf [accessed 2021-12-16]
49. Philipp M, Zeileis A, Strobl C. A toolkit for stability assessment of tree-based learners. Working Papers 2016-11 - Faculty of Economics and Statistics, University of Innsbruck. 2016. URL: https://ideas.repec.org/p/inn/wpaper/2016-11.html [accessed 2021-02-02]
50. Hothorn T, Hornik K, Zeileis A. Unbiased recursive partitioning: a conditional inference framework. J Computat Graph Stat 2006 Sep;15(3):651-674. [doi: 10.1198/106186006x133933]
51. Hothorn T, Zeileis A, Ong C. partykit: A Modular Toolkit for Recursive Partytioning in R. J Mach Learn Res 2015:3905-3909 [FREE Full text]

52. Garnett C, Perski O, Tombor I, West R, Michie S, Brown J. Predictors of engagement, response to follow up, and extent of alcohol reduction in users of a smartphone app (Drink Less): secondary analysis of a factorial randomized controlled trial. JMIR Mhealth Uhealth 2018 Dec 14;6(12):e11175 [FREE Full text] [doi: 10.2196/11175] [Medline: 30552081]
53. Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. Am Psychol 2000;55(1):68-78. [doi: 10.1037/0003-066x.55.1.68]
54. Balmford J, Borland R. How do smokers use a smoking cessation text messaging intervention? Nicotine Tob Res 2014 Dec 16;16(12):1586-1592. [doi: 10.1093/ntr/ntu111] [Medline: 25031312]
55. Heminger C, Boal A, Zumer M, Abroms L. Text2Quit: an analysis of participant engagement in the mobile smoking cessation program. Am J Drug Alcohol Abuse 2016 Jul;42(4):450-458. [doi: 10.3109/00952990.2016.1149591] [Medline: 27120396]

Abbreviations

OR: odds ratio
RF: relative frequency

Edited by R Kukafka; submitted 09.03.21; peer-reviewed by JM Elling, J Keller, D Szinay; comments to author 09.06.21; revised version received 13.09.21; accepted 29.10.21; published 19.01.22.

Please cite as:
Paz Castro R, Haug S, Debelak R, Jakob R, Kowatsch T, Schaub MP
Engagement With a Mobile Phone–Based Life Skills Intervention for Adolescents and Its Association With Participant Characteristics and Outcomes: Tree-Based Analysis
J Med Internet Res 2022;24(1):e28638
URL: https://www.jmir.org/2022/1/e28638
doi:10.2196/28638
PMID:35044309

Original Paper

Behavioral and Self-reported Data Collected From Smartphones for the Assessment of Depressive and Manic Symptoms in Patients With Bipolar Disorder: Prospective Observational Study

Monika Dominiak[1,2*], MD, PhD; Katarzyna Kaczmarek-Majer[3*], PhD; Anna Z Antosik-Wójcińska[2,4], MD, PhD; Karol R Opara[3], PhD; Anna Olwert[3], PhD; Weronika Radziszewska[3], PhD; Olgierd Hryniewicz[3], PhD; Łukasz Święcicki[5], MD, PhD; Marcin Wojnar[4], MD, PhD; Paweł Mierzejewski[1,2], MD, PhD

[1]Department of Pharmacology and Physiology of the Nervous System, Institute of Psychiatry and Neurology, Warsaw, Poland

[2]Section of Biological Psychiatry of the Polish Psychiatric Association, Warsaw, Poland

[3]Department of Stochastic Methods, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

[4]Department of Psychiatry, Medical University of Warsaw, Warsaw, Poland

[5]Department of Affective Disorders, II Psychiatric Clinic, Institute of Psychiatry and Neurology, Warsaw, Poland

[*]these authors contributed equally

Corresponding Author:
Monika Dominiak, MD, PhD
Department of Pharmacology and Physiology of the Nervous System
Institute of Psychiatry and Neurology
Sobieskiego 9
Warsaw, 02-957
Poland
Phone: 48 507183375
Email: mdominia@wp.pl

Abstract

Background: Smartphones allow for real-time monitoring of patients' behavioral activities in a naturalistic setting. These data are suggested as markers for the mental state of patients with bipolar disorder (BD).

Objective: We assessed the relations between data collected from smartphones and the clinically rated depressive and manic symptoms together with the corresponding affective states in patients with BD.

Methods: BDmon, a dedicated mobile app, was developed and installed on patients' smartphones to automatically collect the statistics about their phone calls and text messages as well as their self-assessments of sleep and mood. The final sample for the numerical analyses consisted of 51 eligible patients who participated in at least two psychiatric assessments and used the BDmon app (mean participation time, 208 [SD 132] days). In total, 196 psychiatric assessments were performed using the Hamilton Depression Rating Scale and the Young Mania Rating Scale. Generalized linear mixed-effects models were applied to quantify the strength of the relation between the daily statistics on the behavioral data collected automatically from smartphones and the affective symptoms and mood states in patients with BD.

Results: Objective behavioral data collected from smartphones were found to be related with the BD states as follows: (1) depressed patients tended to make phone calls less frequently than euthymic patients (β=−.064, *P*=.01); (2) the number of incoming answered calls during depression was lower than that during euthymia (β=−.15, *P*=.01) and, concurrently, missed incoming calls were more frequent and increased as depressive symptoms intensified (β=4.431, *P*<.001; β=4.861, *P*<.001, respectively); (3) the fraction of outgoing calls was higher in manic states (β=2.73, *P*=.03); (4) the fraction of missed calls was higher in manic/mixed states as compared to that in the euthymic state (β=3.53, *P*=.01) and positively correlated to the severity of symptoms (β=2.991, *P*=.02); (5) the variability of the duration of the outgoing calls was higher in manic/mixed states (β=.0012, *P*=.045) and positively correlated to the severity of symptoms (β=.0017, *P*=.02); and (6) the number and length of the sent text messages was higher in manic/mixed states as compared to that in the euthymic state (β=.031, *P*=.01; β=.015, *P*=.01; respectively) and positively correlated to the severity of manic symptoms (β=.116, *P*<.001; β=.022, *P*<.001; respectively). We also observed that self-assessment of mood was lower in depressive (β=−1.452, *P*<.001) and higher in manic states (β=.509, *P*<.001).

Conclusions: Smartphone-based behavioral parameters are valid markers for assessing the severity of affective symptoms and discriminating between mood states in patients with BD. This technology opens a way toward early detection of worsening of the mental state and thereby increases the patient's chance of improving in the course of the illness.

(J Med Internet Res 2022;24(1):e28647) doi:10.2196/28647

KEYWORDS

bipolar disorder; generalized linear model; mixed-effects regression; classification; manic episodes; depressive episodes; smartphone; behavioral markers; mHealth; remote monitoring

Introduction

Bipolar disorder (BD) is a chronic, recurrent, and highly morbid illness [1]. Its prevalence is estimated to be around 2%-3% [2]. In the course of the illness, there are fluctuations between different mood states, ranging from depression to hypomanic/manic episodes, as well as mixed states. Patients receiving mood-stabilizing drugs have reported relapse rates of 22% per year [3]. The subsequent episodes seem to worsen the prognosis and increase the suicide risk [4,5]. Patients in the initial phases of BD appear to better respond to treatment; thus, early intervention strategies could be vital for improving illness outcomes [6] by reducing conversion rates to full-blown illness and reducing symptom severity.

Mobile apps allow for real-time collection of both self-reported and objective data on behavioral activities or speech in naturalistic settings [7-12]. Active monitoring through self-reported data was found to correlate with scores on the depression (Hamilton Depression Rating Scale [HDRS]) and mania scales (Young Mania Rating Scale [YMRS]) [7,8,12]. As there are disturbances in diurnal rhythms and daily life regularity of patients with BD [13], continuous self-monitoring might be helpful for managing daily activities. Moreover, self-assessment data can be easily shared with mental health professionals and facilitate making clinical decisions [8,14]. Nevertheless, such monitoring could be less reliable in case of hypomanic/manic symptoms owing to decreased illness insight [12]. The other limitation for certain patients might be the necessity to respond to daily surveys [15]. It has not been also proven so far that such self-monitoring could predict phase change [8].

Data collected through passive objective monitoring of behavioral activities such as phone call statistics, data on physical activity and mobility, or voice features are correlated with scores on the HDRS and YMRS [8-11]. They could serve as markers for monitoring illness activity; thus, behavioral tracking is very promising in recognizing and predicting mood state [12]. Passively collected data do not require any involvement of the patient and could be beneficial in the early detection of subthreshold symptoms of episode recurrence. Disadvantages of passive behavioral tracking include privacy concerns, the discomfort of being observed, or problems related to the collection, analysis, and processing of data. Nevertheless, both active and passive monitoring through smartphone are promising for the assessment of illness activity in patients with BD.

Recent reviews by Antosik-Wójcińska et al [16], Rajagopalan et al [17], and Torous et al [18] agree that the medical potential of smartphone-based monitoring is high, but large-scale methodologically rigorous studies are necessary to draw well-generalizable conclusions. In another review, Rucci et al [19] argue that although nothing can substitute the clinical assessment, smartphone-based monitoring has the potential of improving the treatment owing to its accessibility to patients and the possibility of continuous tracking parameters reflecting illness activity. This could be of great importance, as the frequency of routine assessments during follow-up visits is often insufficient and the latency in identifying the recurrence of a mood episode is often long—3 weeks or more [20]. Consequently, adjustment of the treatment often takes place late—sometimes during a full episode [21]. This is particularly relevant in hypomanic/manic episodes because manic patients, often owing to decreased insight, do not seek medical help [20,22]. Notably, patients with BD are generally open to using smartphones to help them monitor their mental state [23,24]. The usefulness and ease of use of such apps were found to influence patients' satisfaction and adherence [25,26]. Nevertheless, numerous challenges exist, especially concerning safeguarding privacy and ensuring data security [27].

In this paper, we report the results of a prospective observational study of patients with BD. The main goal was to assess whether behavioral data about smartphone usage collected automatically via a dedicated app, called BDmon, correlate with the severity of symptoms on both the HDRS and the YMRS and with the corresponding affective states in BD. Evaluation of the effect size and its statistical significance was based on generalized linear models with random effects for each patient. The secondary objective was to evaluate the suitability, in terms of completeness, of self-assessment data collected via the app over time in naturalistic settings.

Methods

Study Participants

This study was conducted in the Department of Affective Disorders, Institute of Psychiatry and Neurology in Warsaw and in a center specializing in clinical trials between September 2017 and December 2018. Patients were enrolled from both inpatient and outpatient settings to capture both mild and severe episodes. Based on previous studies [28,29], it was expected that patients would change the phase at least once in the period envisaged for this study. This study obtained the consent of the Bioethical Commission at the District Medical Chamber in Warsaw (agreement KB/1094/17).

We used the following inclusion criteria: adults aged 18 years or older, who gave their informed written consent to participate in this study, diagnosed with BD, with at least two changes of phase in the last 12 months, and using a smartphone with internet access daily or declaring their willingness to use it for the study period. Exclusion criteria consisted of serious hearing problems and speech disorders (eg, dysarthria, aphasia).

Sample Size

The resulting sample included 84 eligible patients diagnosed with BD (according to the International Classification of Diseases, tenth revision [ICD-10]). The baseline sociodemographic and clinical characteristics of the final study sample are presented in Table 1.

Table 1. Sociodemographic and clinical characteristics of the final study sample (N=51).

Characteristics	Values
Participation time (days), mean (SD)	208 (32)
Age (years), mean (SD)	36.2 (9.5)
Gender, n (%)	
Female	28 (55)
Male	23 (45)
Demographic living status, n (%)	
Family	30 (59)
Partner	9 (18)
Self	12 (24)
Education, n (%)	
Elementary	2 (4)
Secondary	18 (35)
Higher	31 (61)
Demographic residence, n (%)	
City	37 (73)
Town	11 (22)
Village	3 (6)
Occupation, n (%)	
Working	29 (57)
Pensioner	11 (22)
Student	6 (12)
Other	5 (9)
Clinical characteristics	
Mean duration of illness (years), mean (SD)	7.1 (5.3)
Hospitalizations, median (IQR)	2 (0-7)
Affective episodes, median (IQR)	6 (2-10)
Bipolar disorder type I, n (%)	31 (61)
Bipolar disorder type II, n (%)	20 (39)

Psychiatric Assessment

The assessment of the mental state was carried out by psychiatrists with experience in the diagnosis and treatment of BD. Both the researcher and the patient were blinded to the data automatically collected by the BDmon app. Patients were invited to visit the researcher at least every 3 months. In the case of a suspected change in the mood state, patients were invited for an additional intervention visit. During personal visits, the primary outcome measures (17-point version of HDRS and YMRS) were used. In this paper, the assessment of HDRS and YMRS is used as binding to assess the severity of symptoms and the corresponding BD states.

The data concerning clinical and demographic characteristics were collected as a part of the initial visit. During the following visits, the patients were inquired about significant changes in the general health status and other important life events that might affect their activity and behavior. The patient's mood state was also assessed using fortnightly phone-based interactions with the researcher. The basic and the most

important objective of the phone-based assessment was to determine whether any significant change in the mood state was likely to occur since the previous contact with the patient. If so, the patient was invited for an intervention visit, and a full assessment using HDRS and YMRS was conducted. To shorten the time of the fortnightly calls, the researcher read to the patient questions prepared in the telephone questionnaire form (see Multimedia Appendix 1).

Based on HDRS and YMRS, the mood state was classified into 1 of the 4 phases (depression, hypomania/mania, euthymia, or mixed state). In previous studies, researchers have adopted different cutoff points on the HDRS and YMRS for training classifiers [11,12,30]. High thresholds of 13 points on the HDRS and YMRS [12] increase the validity of the diagnosis, focusing on more severe symptoms. The consequence of this approach might be counting the milder symptoms of a given phase to euthymia. We aimed at the identification of early symptoms of phase change and hence adopted lower cutoff points. The state of euthymia was defined as HDRS<8 and YMRS<6, depression as HDRS≥8 and YMRS<6, hypomania/mania as HDRS<8 and YMRS≥6, and mixed state as HDRS≥8 and YMRS≥6.

Smartphones collect behavioral data continuously whereas clinical assessments are much less frequent. Supervised learning requires both data types to be present. Consequently, we extrapolated the psychiatric assessment to 7 days before and 2 days after a visit. In the final sample, we included only patients with automatically collected smartphone data within this period. The asymmetry of the time window was inspired by Muaremi et al [31] who argue that a visit is typically related to some medical intervention, which might lead to a change in the patient's mental state. The time frame before the visit differs across studies, for example, Faurholt-Jepsen et al [11,12] adopted a time span of the clinical assessment days and 3 previous days. In this study, we extended this period up to 7 days before the visit because we had additional data about patient's mental state (derived from phone-based visit and from patients' relatives). Nevertheless, to facilitate meta-analyses, we additionally performed the analysis for the shorter period assuming ground truth of 3 days before the psychiatric assessment and the day of psychiatric assessment, as well for different cutoff points (Multimedia Appendix 2 and Multimedia Appendix 3).

Smartphone-Based Data Collection

The BDmon app requires a smartphone with the Android system, which at the time of the clinical trial had over 90% of the market share in Poland. Patients used their own smartphones or were offered a new smartphone for the study period. Patients did not receive any financial gratification for participation in the study. The BDmon app collected 3 groups of data:

(1) objective data about patient's behavior related to smartphone (statistics about phone calls and text messages), (2) self-reported data completed by patients via the app (filling in the questionnaires was not mandatory; Multimedia Appendix 4), and (3) objective acoustic features about patient's speech extracted from his/her daily phone calls. Only self-assessment (2) required the patients to interact with the software; data from groups (1) and (3) were collected passively. The scope of monitoring was as follows:

1. Objective data about patient's behavior consisted of statistics of calls and text messages transformed into 11 daily aggregates describing behavioral parameters. They fall into 4 categories: (1) incoming answered calls: number (per day), mean duration (seconds/call), and variability of duration (the standard deviation of the lengths of all calls/day; seconds); (2) incoming missed calls: number (per day) and fraction of missed calls; (3) outgoing calls: number (per day), mean duration (seconds/call), variability of duration (standard deviation/day; seconds), and fraction of outgoing calls (ratio of the number of outgoing calls to the sum of outgoing and incoming calls); and (4) outgoing text messages: number (per day) and average number of characters in outgoing messages (per day). Initially, the app collected data concerning mobility (daily travelled distance) and activity (pedometer), but owing to the technical issues and patients' privacy concerns, these data were mostly missing and were not eventually analyzed.
2. Self-reported data completed by patients via the app: filling in the self-assessment questionnaires was not mandatory to evaluate, in natural settings, their completeness over time, and relevance for analysis. Patients assessed their well-being using a dedicated mood rating scale created for the purposes of the study. Each patient had access to a graph illustrating the mood status and the length of sleep.
3. Acoustic features of patient's speech signal consisting of 85 physical parameters describing each phone call: acoustic features require dedicated processing, which is beyond the scope of this paper. The preliminary analysis of the acoustic data is a subject of other ongoing and completed works [32,33].

Statistical Methods

To assess the strength of the relation between the behavioral markers and the affective symptoms and BD states, we applied generalized linear mixed-effects models similarly as in the work of Faurholt-Jepsen et al [12]. First, mixed-effects linear regression was applied with scores on either HDRS or YMRS as response variables. This quantifies the correlations between behavioral markers and the severity of affective symptoms. Second, mixed-effects logistic regression was used for binary classification discriminating between an affective state and euthymia to describe the relation between the behavioral markers and affective states. The generalized linear mixed-effects model had the following form:

$$y = X\beta + Zu + \varepsilon$$

where y is an $n_d \times 1$ vector consisting of n_d responses, each of which corresponds to 1 patient and 1 day. When assessing the severity of symptoms, either the HDRS or YMRS scores are used as the response vectors.

X is an $n_d \times s$ covariate matrix of the s predictor variables for fixed effects β (regression coefficient). The fixed effects consist of coefficients related to the objective smartphone data (eg, number of incoming calls) and a constant term. We build a

separate model for each of the predictor variables; therefore, s=1.

Z is the $n_d \times q$ covariate matrix for the random effects u. The random effects occur at the patient level (level two) and are patient-specific random intercepts. Consequently, q is equal to the number of patients (groups) for whom data were available.

ε is the vector of n_d errors, which is assumed to be multivariate normal with zero mean. The number of patients differs depending on the considered parameter. For example, if patient A was never sending text messages, the daily number of sent text messages was calculated as 0. This allowed us to include patient A for the assessment of the relation between symptoms and the number of sent text messages. At the same time, sending no text messages led to missing values reported for the mean number of characters of sent text messages, and thus, patient A was not counted for assessment of the relation between symptoms and the length of sent text messages. Logistic mixed-effects regression models were applied to assess the relations between pairs of euthymia and the affective states (eg, euthymia vs depression). For each of the logistic models, the response variable equaled 0 for days with the euthymic state and 1 for days with the affective state (eg, depression). The inverse of the log link binomial function $h(\eta) = (1+\exp(\eta))^{-1}$ was applied to the linear predictor $\eta = X\beta + Zu$ to relate it to the outcome y. The mixed-effects logistic regression model was formulated as follows:

$$y = h(X\beta + Zu) + \varepsilon$$

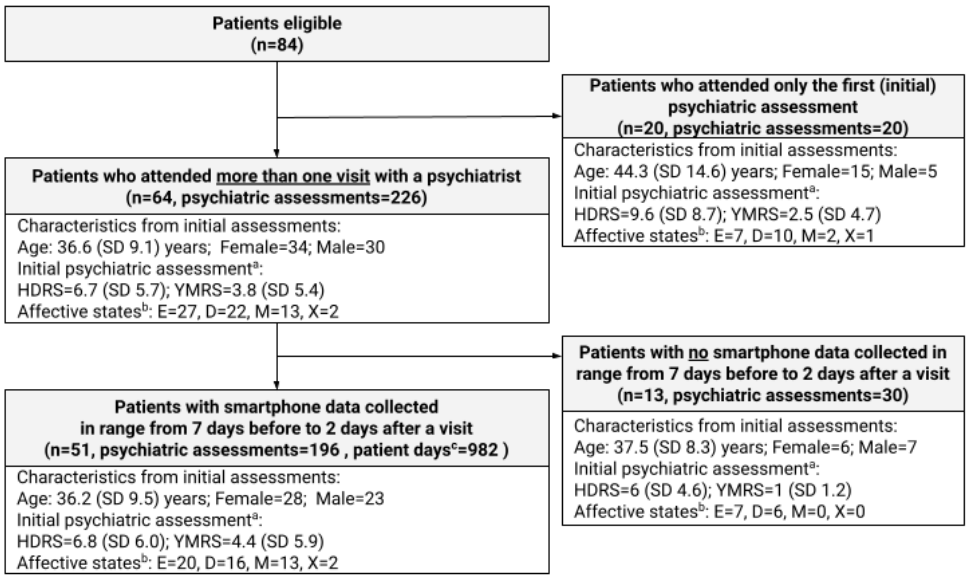
Model assumptions were checked with residual analysis and visually using quantile-quantile plots. The P values were calculated assuming normal distributions of errors. However, for some variables, especially the number of outgoing and incoming calls per day, this assumption seemed to be violated. Therefore, caution is necessary when interpreting the P values of the mixed-effects models; see also Luke [34]. Log transformations were applied to minimize this effect and the results were gathered for comparative purposes in Multimedia Appendix 5 and Multimedia Appendix 6. All analyses were conducted in the R programming language (R Core Team and the R Foundation for Statistical Computing). Linear and logistic mixed-effects models were calculated using packages lme (for the restricted maximum likelihood estimation) and lmerTest (P values and model diagnostics) available at the Comprehensive R Archive Network repository for R language [35]. We set the significance level to .05.

Results

Study Sample

Figure 1 illustrates the participant flow in this study; 84 eligible patients diagnosed with BD (according to ICD-10 classification) were enrolled in this study, and they participated in the initial interview with the psychiatrist. Disregarding their declarations of participation in this study, 20 patients dropped out after the initial psychiatric assessment. As depicted in Figure 1, the mean age (36.6 [SD 9.1] years) was lower for the group of 64 patients who continued the study in comparison to that of the group of 20 patients who resigned after the initial visit (44.3 [SD 14.6] years). For the remaining 64 patients, there were 226 psychiatric assessments in total. However, smartphone-based automatically collected behavioral data in the assumed time frame (7 days before and 2 days after psychiatric assessment) were available for only 51 patients (196 psychiatric assessments). These patients (N=51) were considered as the final study sample. In total, this constitutes 982 patient days with data for the statistical analyses. The mean participation time for the final sample of 51 patients, calculated as the difference between the initial and the last psychiatric assessment, was 208 days with a standard deviation of 132 days.

Figure 1. Flow chart illustrating the number of patients, psychiatric assessments, and patient days during the study.



[a]Mean (SD) of Hamilton Depression Rating Scale (HDRS) and Young Mania Young Mania Rating Scale (YMRS) from initial psychiatric assessments.
[b]Euthymia (E) was defined as HDRS < 8 and YMRS < 6, depression (D) as HDRS ≥ 8 and YMRS < 6, hypomania/mania (M) as HDRS < 8 and YMRS ≥ 6, mixed state (X): HDRS ≥ 8 and YMRS ≥ 6.
[c]Ground truth for the analyses is 7 days before to 2 days after a visit. A patient day is counted if the BDmon app collected data from at least a single SMS or a single call.

Psychiatric Assessments and Affective States

Overall, 196 psychiatric assessments were reported for the final study sample. We observed 145 mood state transitions among consecutive psychiatric interviews as summarized in Table 2.

Out of 145 transitions, in 74 (51%) cases, the affective state changed from the previous one and in 71 (48.9%) remained the same. The most frequent change was from euthymia to depression state (17 such cases) and from depression to euthymia (14 cases).

Table 2. Transitions among the affective states for consecutive visits for patients of this study (N=51).

Flow from the following phases	To the following phases			
	Euthymia	Depression	Mixed state	Mania
Euthymia	25	17	2	6
Depression	14	39	4	4
Mixed state	4	2	1	2
Mania	9	5	5	6

Completeness of Data Collected From Smartphones

Figure 2 illustrates the completeness of the objective behavioral data collected from smartphones. For each patient, there are 2 rows of dots. Green dots in the upper row show days, in which the objective behavioral data were collected. Turquoise dots in the lower row indicate days with self-assessment data. Less frequent usage of the app in early 2018 was a result of technical stability issues with the app.

Overall, we identified 982 person days with smartphone data and psychiatric assessments. The summary statistics of behavioral and self-assessment data are presented in Table 3 for all variables as initially planned for this study.

Figure 2. Completeness of data: for each patient, the green dots mean that behavioral data were collected, turquoise dots represent self-assessment data, and letters denote the clinically assessed affective state of patients with bipolar disorder. D: depression; E: euthymia; M: mania; X: mixed.

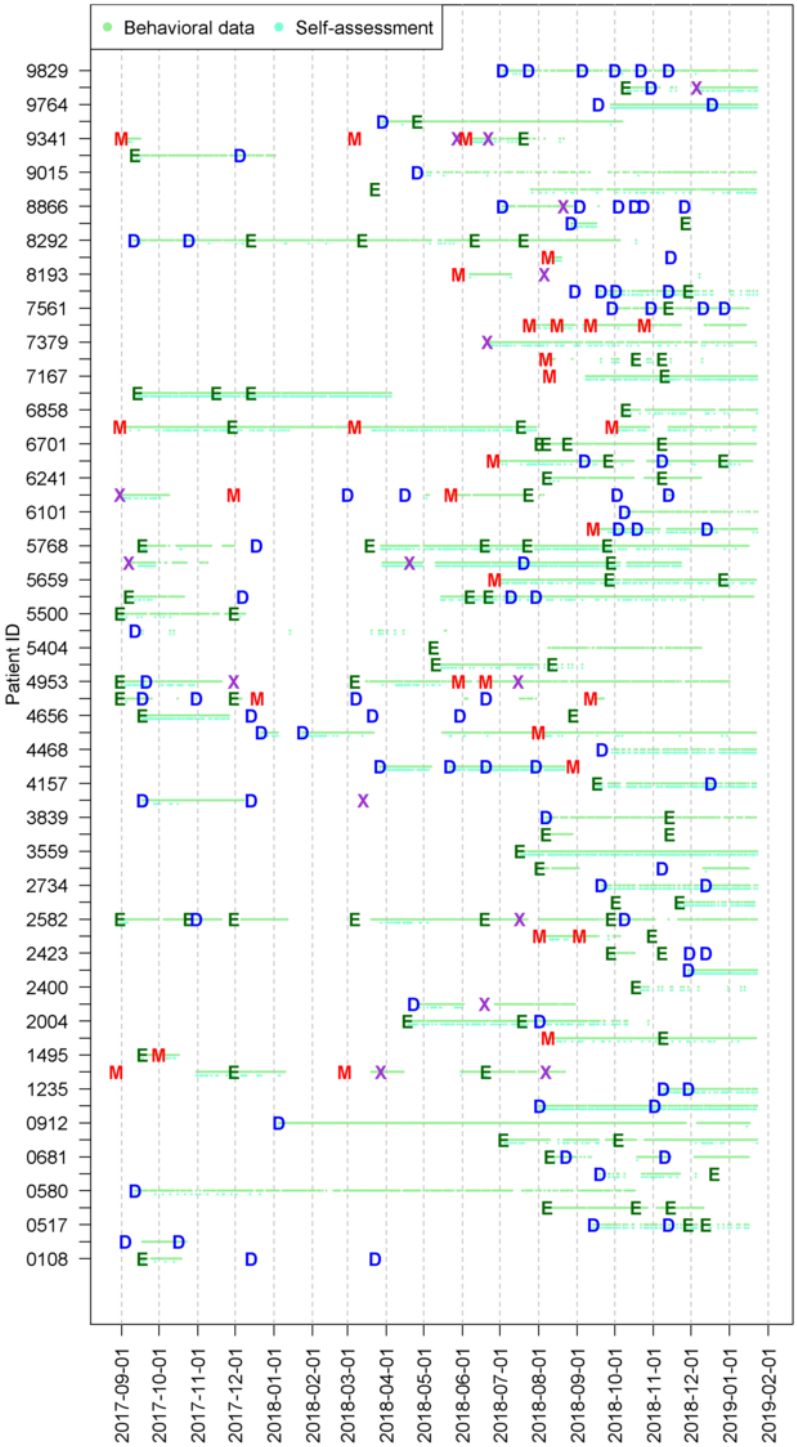


Table 3. Basic summary statistics of the behavioral and self-assessment data collected from smartphones.

Type, daily variable	Completeness, n (%)	Median	Mean (SD)
Phone calls (count per patient day)			
Number of incoming answered calls	982 (100)	2.0	3.2 (3.5)
Mean duration of incoming calls (s/call)	820 (83.5)	119.9	210.3 (286.3)
SD of duration of incoming calls (s)	600 (61.1)	117.7	234.5 (324.3)
Number of outgoing calls	982 (100)	4.0	7.5 (10.9)
Mean duration of outgoing calls (s/call)	886 (90.2)	79.1	147.2 (232.8)
SD of duration of outgoing calls (s)	756 (77.4)	113.3	195.9 (245.7)
Fraction of outgoing calls	886 (90.2)	0.7	0.7 (0.2)
Number of missed calls	982 (100)	1.0	1.8 (2.7)
Fraction of missed calls	982 (100)	0.2	0.2 (0.2)
Number of incoming + outgoing calls	963 (98.0)	105.8	170.2 (229.4)
Mean duration of incoming + outgoing calls (s/call)	877 (89.3)	145.8	234.0 (256.8)
SD of duration of incoming + outgoing calls (s)	982 (100)	8.0	12.5 (14.4)
Text messages			
Number of sent text messages	982 (100)	0.0	3.8 (13.3)
Mean length of text messages	343 (34.9)	39.4	54.3 (48.7)
Activity			
Sum of steps	103 (10.5)	480	5299.9 (32454.8)
Daily travelled distance (km)	__[a]	—	—
Self-assessment			
Self-assessment of sleep time (h)	270 (27.5)	8.0	8.0 (2.5)
Self-assessment of mood (from −4 to +4)	268 (27.3)	0.0	−0.6 (1.6)
Comment about sleep (number of characters)	39 (4)	22.0	27.1 (21.0)
Comment about mood (number of characters)	37 (4)	25.0	40.2 (46.6)

[a]Not collected due to technical issues and privacy concerns.

The app was designed to collect data concerning mobility (daily travelled distance) and the activity measures with the number of steps (pedometer), but owing to the technical issues and patients' privacy concerns, the completeness of these data (98/982, 9.9%) was insufficient to conduct a reliable analysis. Similarly, the collected self-reported comments about sleep and mood were characterized by relatively high rates of missing data (n=39 and n=37, respectively). Table 4 provides the sociodemographic and clinical characteristics of patients depending on their adherence in terms of filling in self-assessment questionnaires. The study patients were split into 3 groups depending on the completeness of the self-assessment data; 16 patients out of 51 patients (31%) demonstrated low adherence to filling in self-assessment questionnaires (completeness less than 98 patient days). The group of patients with low adherence had relatively high sum of points on the HDRS (7.25 [SD 5.05]) and low sum of points on the YMRS (1.81 [SD 1.64]). Interestingly, 10 patients out of 16 in the group with low adherence (62.5%) were assessed as euthymic during the initial assessment.

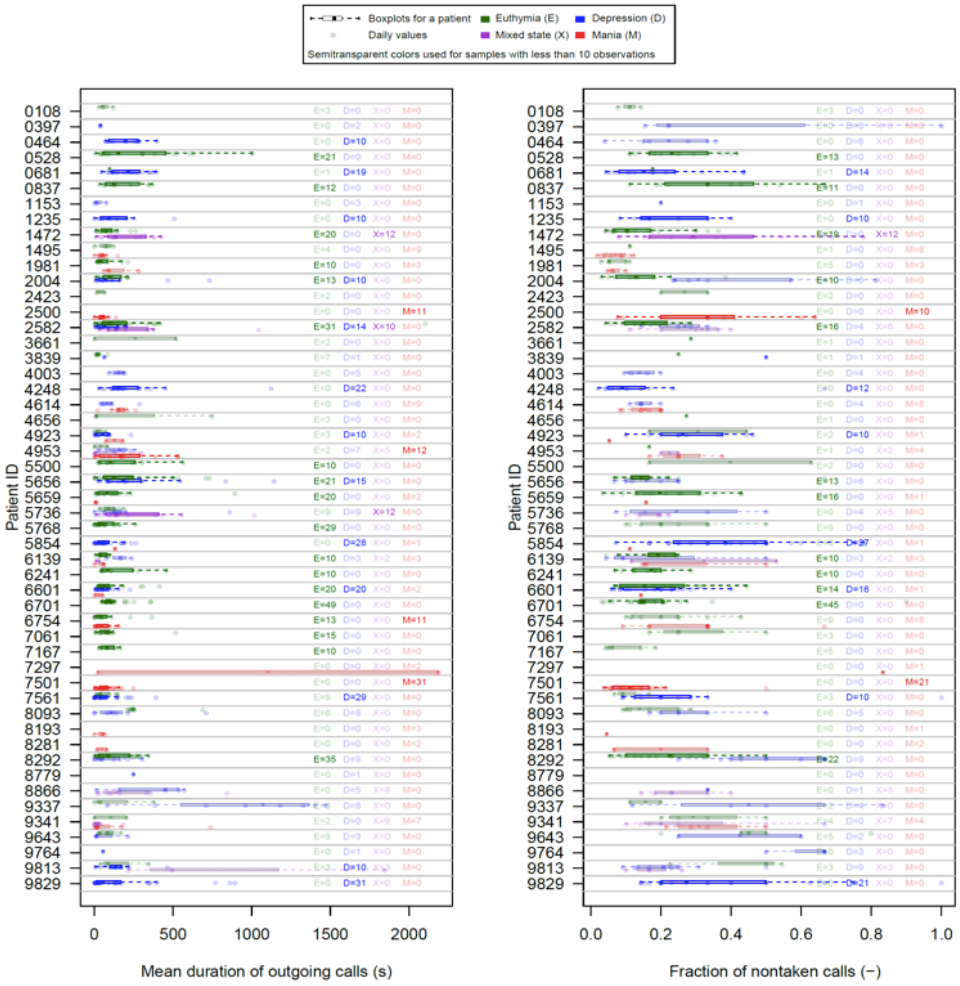
Table 4. Sociodemographic and clinical characteristics of the patients grouped according to their adherence of filling the self-assessment questionnaires (N=51).

Characteristics	Data completeness of filling the self-assessment		
	Low (<10%)	Medium (10%-90%)	High (>90%)
Group of patients according to adherence of filling the self-assessment, n (%)	16 (31)	29 (57)	6 (12)
Gender, n (%)			
Female	7 (14)	18 (35)	3 (6)
Male	9 (18)	11 (22)	3 (6)
Demographic living status, n (%)			
Family	11 (22)	13 (25)	6 (12)
Partner	2 (4)	7 (14)	0
Self	3 (6)	9 (18)	0
Education, n (%)			
Elementary	0	2 (4)	0
Secondary	4 (8)	10 (20)	4 (8)
Higher	12 (24)	17 (33)	2 (4)
Demographic residence, n (%)			
City	11 (22)	22 (43)	4 (8)
Town	4 (8)	6 (12)	1 (2)
Village	1 (2)	1 (2)	1 (2)
Age (years), mean (SD)	37.31 (8.66)	36.41 (8.69)	32.67 (15.28)
Severity of symptoms from the initial psychiatric assessment, mean (SD)			
Hamilton Depression Rating Scale	7.25 (5.02)	6.97 (6.89)	5 (4.15)
Young Mania Rating Scale	1.81 (1.64)	5.48 (6.35)	6.5 (8.65)
Affective state assessed during initial psychiatric assessment (n)			
Euthymia	10	7	3
Depression	6	9	1
Mania	0	11	2
Mixed state	0	2	0

Between-Patient and Intrapatient Variability

The summary statistics about the behavioral and self-assessment data in various affective states are shown in Multimedia Appendix 7. For most of the considered variables, differences in statistics were observed between the 4 groups (euthymia, depression, mania, mixed state). Apart from the overall difference of statistics for patients in various BD states, there was also a relatively high intrapatient variability in phone call statistics. Figure 3 illustrates this for 2 objective parameters and all patients with BD. The boxplots for all the variables are presented in Multimedia Appendices 8-19.

Figure 3. Intrapatient variability of 2 objective parameters collected with smartphones among chosen patients with bipolar disorder.



Relation Between Smartphone-Based Data and Severity of Depressive and Manic Symptoms

Table 5 presents the estimates of the regression coefficients β. They describe the influence of behavioral data on the severity of manic and depressive symptoms measured by the number of HDRS and YMRS points. The number of observations is the product of the number of groups (ie, patients) and days for which the considered data type was collected.

The results show that the more severe the manic symptoms: (1) the higher variability in the duration of incoming calls (P=.04) and outgoing calls (P=.02), (2) the higher number and fraction of missed calls (P=.02), (3) the higher number of outgoing text messages (P<.001), (4) the higher number of characters in outgoing text messages (P<.001), and (5) the higher score on self-reported mood rating (P<.001). For example, for every increase of 1 missed call, there was an increase of 0.141 points on the YMRS. More severe depressive symptoms are associated with (1) lower number of incoming answered calls (P<.001), (2) higher fraction of incoming missed calls (P<.001), and (3) lower score on self-reported mood rating (P<.001).

Table 5. Relations between smartphone-based data collected using the BDmon app and depressive and manic symptoms assessed with the Hamilton Depression Rating Scale and Young Mania Rating Scale, respectively.

Daily variable	Depressive symptoms to			Manic symptoms			Patient days (n)	Patients (n)
	Regression coefficient (β)	*P* value	95% CI	Regression coefficient (β)	*P* value	95% CI		
Number of incoming answered calls	−.149	<.001	−0.231 to −0.066	−.079	.10	−0.174 to 0.016	982	51
Duration of incoming calls (s/call)	.001	.17	0 to 0.002	.001	.22	0 to 0.002	820	50
Standard deviation of incoming calls duration (s)	.001	.29	0 to 0.001	.0012	.04	0 to 0.002	600	48
Number of outgoing calls	−.023	.09	−0.05 to 0.004	.002	.90	−0.029 to 0.033	982	51
Fraction of outgoing calls	.307	.64	−0.994 to 1.604	.985	.19	−0.492 to 2.471	886	51
Duration of outgoing calls (s/call)	.001	.37	−0.001 to 0.002	.001	.09	0 to 0.002	886	51
Standard deviation of outgoing calls duration (s)	0	.91	−0.001 to 0.001	.0017	.02	0 to 0.003	756	48
Number of missed calls	.06	.27	−0.047 to 0.167	.141	.02	0.019 to 0.263	982	51
Fraction of missed calls	4.861	<.001	2.829 to 6.896	2.991	.02	0.434 to 5.526	582	50
Number of sent text messages	.02	.06	−0.001 to 0.041	.116	<.001	0.093 to 0.139	982	51
Mean length of text messages (number of characters)	.004	.30	−0.003 to 0.011	.022	<.001	0.011 to 0.034	343	38
Self-assessment of sleep time (h)	−.027	.83	−0.274 to 0.22	−.130	.19	−0.325 to 0.062	270	39
Self-assessment of mood	−1.452	<.001	−1.777 to −1.125	.509	<.001	0.267 to 0.751	268	42

Relation Between Smartphone Data and Affective States

Tables 6 and 7 present the results from generalized mixed regression models (logistic model). Smartphone-based data were used as predictors. We were interested in the detection of behavioral changes between euthymia, which we treated as a reference and encode as 0, and the other affective states encoded as 1. Therefore, for each predictor, we fitted 4 regression models distinguishing euthymia from depression, mania, mixed state as well as a combination of manic and mixed states. Positive values of the regression coefficients indicate that the predictor tends to have lower value in the reference euthymic state. As we put emphasis on the early identification of phase change by adapting lower cutoff points on the above scales, we were able to distinguish between euthymia and other affective states, both mild and severe, that is, depression, hypomania/mania, or mixed state.

Table 6. Mixed regression models regarding behavioral smartphone-based data and affective states (euthymia, depression, and mania) in patients with bipolar disorder assessed with the Hamilton Depression Rating Scale and Young Mania Rating Scale.

Daily variable	Euthymia versus depression					Euthymia versus mania				
	Regression coefficient (β)	*P* value	95% CI	Observations (n)	Patients (n)	Regression coefficient (β)	*P* value	95% CI	Patient days (n)	Patients (n)
Number of incoming answered calls	−.15	.01	−0.243 to −0.056	789	46	.024	.71	−0.103 to 0.152	567	41
Duration of incoming calls (s/call)	0	.41	0 to 0.001	664	46	−.001	.59	−0.004 to 0.002	488	39
Standard deviation of duration of incoming calls (s)	0	.45	−0.001 to 0.001	490	44	−.001	.41	−0.005 to 0.002	375	36
Number of outgoing calls	−.064	.01	−0.113 to −0.014	789	46	.044	.16	−0.018 to 0.106	567	41
Fraction of outgoing calls	1	.09	−0.151 to 2.152	715	46	2.73	.03	0.223 to 5.237	526	41
Duration of outgoing calls (s)	.001	.31	−0.001 to 0.002	715	46	0	.88	−0.003 to 0.003	526	41
Standard deviation of duration of outgoing calls (s)	0	.95	−0.001 to 0.001	603	42	−.002	.31	−0.006 to 0.002	463	41
Number of missed calls	−.017	.77	−0.132 to 0.098	789	46	−.03	.76	−0.227 to 0.167	567	41
Fraction of missed calls	4.431	<.001	2.12 to 6.742	463	45	1.912	.38	−2.335 to 6.158	345	40
Number of sent text messages	.001	.94	−0.036 to 0.039	789	46	.015	.69	−0.06 to 0.09	567	41
Mean length of text messages (number of characters)	.001	.91	−0.016 to 0.018	265	28	.014	.08	−0.002 to 0.029	192	28
Self-assessment of sleep time (h)	−.254	.08	−0.538 to 0.03	219	33	−.548	.19	−1.369 to 0.273	147	30
Self-assessment of mood	−2.056	<.001	−2.058 to −2.055	224	35	.676	.22	−0.406 to 1.758	144	31

Table 7. Mixed regression models regarding smartphone-based data and affective states (euthymia, mixed/manic states) in patients with bipolar disorder assessed with the Hamilton Depression Rating Scale and Young Mania Rating Scale.

Daily variable	Euthymia versus mixed state					Euthymia versus mania and mixed states				
	Regression coefficient (β)	*P* value	95% CI	Observations (n)	Patients (n)	Regression coefficient (β)	*P* value	95% CI	Patient days (n)	Patients (n)
Number of incoming answered calls	−.114	.08	−0.244 to 0.015	516	35	−.043	.32	−0.127 to 0.042	638	42
Duration of incoming calls (s/call)	.002	.06	0 to 0.004	448	34	.001	.26	−0.001 to 0.002	546	40
Standard deviation of duration of incoming calls (s)	.0027	.02	0 to 0.005	339	33	.001	.11	0 to 0.003	412	38
Number of outgoing calls	−.035	.13	−0.081 to 0.01	516	35	.005	.64	−0.016 to 0.026	638	42
Fraction of outgoing calls	.44	.70	−1.799 to 2.68	477	35	1.363	.11	−0.318 to 3.045	587	42
Duration of outgoing calls (s)	.0015	.045	0 to 0.003	477	35	.001	.06	0 to 0.002	587	42
Standard deviation of duration of outgoing calls (s)	.0031	.01	0.001 to 0.005	420	35	.0012	.045	0 to 0.002	518	42
Number of missed calls	.135	.07	−0.012 to 0.282	516	35	.08	.18	−0.036 to 0.195	638	42
Fraction of missed calls	4.928	<.001	4.926 to 4.93	310	35	3.53	.01	0.907 to 6.153	387	42
Number of sent text messages	.032	.01	0.009 to 0.055	516	35	.031	.01	0.009 to 0.052	638	42
Mean length of text messages (number of characters)	.014	.07	−0.001 to 0.029	180	21	.015	.01	0.003 to 0.026	225	29
Self-assessment of sleep time (h)	−.121	.62	−0.606 to 0.363	144	24	−.334	.08	−0.708 to 0.04	171	31
Self-assessment of mood	.166	.69	−0.666 to 0.998	142	25	.348	.31	−0.327 to 1.022	165	32

The following variables discriminate between euthymia and depression: (1) number of incoming answered calls (*P*=.01), (2) fraction of missed calls (*P*<.001), and (3) number of outgoing calls (*P*=.01). Euthymia and mania differ significantly in fraction of outgoing calls (*P*=.03). The mixed states observed in this study were mainly mixed manic states, which was reflected in the average scores on HDRS (10.6 [SD 2.9]) and YMRS (17.6 [SD 10.2]). This is the rationale behind conducting additional analysis to catch the whole spectrum of manic or mixed features by combining both states into 1 group in Table 7. The following variables turned out to be relevant and discriminate the above conditions from euthymia: (1) fraction of missed calls (*P*=.01), (2) variability of the duration of calls (*P*=.045), (3) number of sent text messages (*P*=.01), and (4) mean length of text messages (*P*=.01).

Further, in Table 8, we discriminate euthymia from the pathological states considered as 1 group to find out which predictors are relevant markers for BD. The following 7 out of 13 variables (54%) are statistically significant: (1) number of incoming calls (*P*=.002), (2) fraction of outgoing calls (*P*=.01), (3) fraction of missed calls (*P*<.001), (4) number of sent text messages (*P*=.01), (5) mean length of text messages (*P*=.01), (6) self-reported sleep time (*P*=.03), and (7) self-reported mood (*P*=.003). It needs to be noted that 6 out of these 7 variables (all except for the self-reported sleep time) are statistically significant and discriminate euthymia from the pathological states considered as 1 group also for the longer period (14 days) preceding the psychiatric assessments (see Multimedia Appendices 8-20).

Table 8. Mixed regression models regarding smartphone-based data and affective states (euthymia vs depressive/mixed/manic states) in patients with bipolar disorder assessed with the Hamilton Depression Rating Scale and Young Mania Rating Scale.

Daily variable	Euthymia versus depression/mania/mixed state				
	Regression coefficient (β)	*P* value	95% CI	Observations (n)	Patients (n)
Number of incoming answered calls	–.101	.002	–0.164 to –0.038	982	51
Duration of incoming calls (s/call)	.000	.29	0 to 0.001	820	50
Standard deviation of duration of incoming calls (s)	.001	.24	0 to 0.001	600	48
Number of outgoing calls	–.005	.55	–0.022 to 0.012	982	51
Fraction of outgoing calls	1.211	.01	0.244 to 2.179	886	51
Duration of outgoing calls (s)	.001	.07	0 to 0.002	886	51
Standard deviation of duration of outgoing calls (s)	.0004	.34	0 to 0.001	756	48
Number of missed calls	.0496	.24	–0.033 to 0.132	982	51
Fraction of missed calls	4.494	<.001	4.492 to 4.496	582	50
Number of sent text messages	.023	.01	0.006 to 0.041	982	51
Mean length of text messages (number of characters)	.010	.04	0 to 0.02	343	38
Self-assessment of sleep time (h)	–.261	.03	–0.493 to –0.03	270	39
Self-assessment of mood	–.551	.003	–0.913 to –0.189	268	42

Summary of the Analyses

The most clinically relevant data that could be drawn from both presented analysis types are as follows: (1) number of incoming answered calls was lower in patients with depression as compared to those with euthymia and, at the same time, missed incoming calls were more frequent and increased as depressive symptoms intensified; (2) depressed patients tended to make phone calls less frequently than euthymic patients; (3) fraction of missed calls was higher in manic/mixed states and was positively correlated to manic symptoms; (4) fraction of outgoing calls was higher in manic states; (5) variability of duration of calls was higher in manic/mixed states and positively correlated to the severity of symptoms; and (6) the number and length of sent text messages was higher in manic/mixed states as compared to euthymic state and positively correlated to the severity of manic symptoms. The self-reported mood scores were significantly correlated with depressive and manic symptoms (as measured with the HDRS and YMRS), but data were insufficient to explain their relation to BD states.

Discussion

Principal Findings

To the best of our knowledge, this is one of the largest studies investigating the relation between the BD phase assessed by psychiatrists and the objective behavioral data collected via smartphones. Reliable smartphone-based results were obtained for 51 patients. This paper confirms and further extends the findings presented by Faurholt-Jepsen et al [12] concerning objective behavioral markers based on phone call statistics. Our results confirm that the majority of phone call statistics are significantly correlated with the clinically rated depressive and manic symptoms (assessed using the HDRS and YMRS, respectively) and are valid markers distinguishing between BD phases (mania, depression, mixed state, euthymia). We observed that depressed patients made phone calls less frequently than euthymic patients; they also answered the phone less often, and this behavior increased as the depressive symptoms intensified. However, patients in the manic/mixed states used phones more frequently than patients in the euthymic state, which was seen in the higher fraction of outgoing calls and a higher number and length of text messages. In general, these results are consistent with those reported in a previous research. For example, in the studies of Faurholt-Jepsen et al [8] and Muaremi et al [31], the number and length of phone calls were correlated with the BD phase and were found to be higher in mania and lower in depression. However, the available studies are not entirely homogeneous in terms of the given phone call parameter and its direction of changes in the particular phase of BD [8,12,31,36]. Patients present different behavioral patterns depending on individual characteristics and preferences for using a mobile phone. We suppose that this issue may at least partly depend on the analyzed group of patients and intrapatient variability. In our study, we also observed that patients in manic/mixed states demonstrated various phone usage patterns—from very long calls to very rare phone use. This was especially reflected in the large standard deviation of the duration of calls. To our knowledge, this variability has not been investigated in detail in similar studies so far. This could open up an interesting field for further research focused on tailored solutions for the individual patient rather than on the characteristics of BD itself. It seems that taking precautions in considering call statistics as a single-phase marker or phase change marker for all patients, in particular, in manic/mixed state, would be reasonable. The behavioral parameters differ between mood phases (depression, mania, mixed state) as compared to euthymia. Depression and mania reflect various manifestations of the illness, but not necessarily should be seen as two poles of the same dimension [37]. Consequently,

parameters characterizing depression and mania do not need to have opposite signs but will rather be different from euthymic state. In our study, this conclusion was also confirmed by the observation of missed calls among patients with depression and mania or mixed state. This parameter was significantly increased in both affective states as compared to euthymia and was positively correlated to the severity of depressive and manic symptoms. This could indicate that both manifestations of the disease, although of different affective tones, have a similar effect on social interactions, and its size increases with the severity of the illness symptoms. In depression, this could be due to psychomotor retardation and social isolation, while in manic/mixed states, this could be the effect of growing distractibility and inattention. Therefore, we conjecture that the key information might be conveyed by the general change in behavioral patterns appropriate for a given patient in euthymia, and not the direction of the change itself. The above conclusion is also supported by an analysis comparing all pathological affective states in BD to the euthymic state (Table 8), showing a significant difference with the euthymic state. The correlation of self-assessment data with the clinically rated depressive or manic symptoms has been shown in the previous research [8,38-41]. It is worth noting at this point that there are also apps providing personalized psychoeducation programs or mobile therapy combined with self-assessment tools, for instance, the SIMPLe app [42]. The SIMPLe app showed efficacy in improving sleep, social rhythms, and eating pattern [43].This paper also revealed that in natural settings, patients' adherence declined over time. We observed that at the very beginning, almost 90% of the patients filled data regularly, but after 3 months, only less than one-quarter of the patients systematically continued the self-assessment. This adherence rate appeared to be similar to that reported by Hidalgo-Mazzei et al [44] where only 30% of the patients were using the app regularly after 6 months of the study. Nevertheless, it should be stressed that adherence in completing self-rated questionnaires ranging from 42% to 95% was observed in previous studies [45]. We have identified that most of the patients with low adherence in filling self-assessment questionnaires were in the state of euthymia during the initial psychiatric assessment. The other hypothesis is that cognitive performance might play a certain role in filling the assessments [46]; yet, the data collected in this study provides little information on this topic. Further research might take this potential variable into consideration. The long-term collection of self-assessment data and their suitability for the recognition and prediction of affective state seem to be problematic. Similar conclusions were also drawn by Faurholt-Jepsen et al [8]. The completeness of the self-assessment data collected in this study was insufficient to fully explain their relation to BD states. A strength of this study is that most patients were assessed several times using a longitudinal design for an average of ~7 months. The adopted cutoff points on the depressive (HDRS) and manic (YMRS) scales allowed us to capture changes in phone usage patterns that are already different in mild depression/hypomania/mania/mixed states as compared to euthymia. Patients with severe manic/mixed symptoms also tended to provide more diversified patterns of behavior in the same condition. This sheds new light on, so far little studied, the variability of behavior among patients with BD (measured for example with the standard deviation) in given affective states. Further research could concentrate on personalized apps, adapting to a given patient, or the search for more generalizable smartphone-based objective parameters, which will be independent of the individual patterns of the phone usage and behavior. Acoustic features of the human voice seem to be promising candidates [47]. In the study of Faurholt-Jepsen et al [11], voice data turned out to be a reliable objective phase marker in BD. Moreover, speech features seem to be a promising marker for the assessment of suicide risk in patients with depression [48]. In our study, we also collected these data, but owing to their extensive nature and the need for thorough analysis, they will be presented in a separate paper. Relations between data collected from smartphones and the depressive and manic affective symptoms confirmed in this study demonstrate promising potential for both early detection of affective states and the prediction of phase changes. This is commonly formulated as either classification or regression task [49]. Recent papers [34,50] show that statistical and machine learning approaches can be complemented by process monitoring with control charts. These are easily understandable visual tools that naturally address the temporal structure of data and generate notifications about the change of BD phase.

Limitations

The initially planned sample size was 100 patients. The reason for not meeting this goal was a slowdown in the trial in early 2018 due to technical issues (eg, monitoring of the daily travelled distance based on satellite navigation data, disruption of the normal usage of smartphone). However, it is worth noting that the final sample size (N=51) is still one of the largest among similar studies. The BDmon app was developed only for Android as it was the dominating operating system in Poland at the time of this study and we did not want to add additional costs and technical complexity to the project. During the enrolment into the study, only 3 out of 84 patients (4%) did not have an Android smartphone and were offered a relevant device for the study period. Therefore, we believe that the potential autoselection of our sample arising from this issue was negligible. The most controversial monitoring function for patients was travelled distance computed using satellite navigation—almost 90% of the patients refused to be monitored; thus, these data were not eventually analyzed. Moreover, patients in the manic state were likely to switch off the smartphone quite frequently or uninstall the app, which created nonrandom missing data. During this study, 3 patients experienced worsening of the psychotic symptoms accompanying depressive and manic episodes. Finally, since this study was exploratory, no correction for multiple comparisons was applied.

Conclusions

This study brings strong evidence that smartphone-based parameters reflecting behavioral activities are related to the severity of depressive and manic symptoms and allow for discriminating between affective states in BD, that is, depression versus euthymia and manic/mixed states versus euthymia. These parameters could be used to assess the severity of manic or depressive symptoms in BD and assist in the early recognition

of phase change, which can increase the patient's chance of early intervention between outpatient visits and hence improve the course of the disease and prognosis.

Acknowledgments

This project was supported by EU funds (Regional Operational Program for Mazovia), a project entitled "Smartphone-based diagnostics of phase changes in the course of bipolar disorder" (RPMA.01.02.00-14-5706/16-00).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Telephone questionnaire form.
[PDF File (Adobe PDF File), 615 KB - jmir_v24i1e28647_app1.pdf]

Multimedia Appendix 2
Changed ground truth.
[PDF File (Adobe PDF File), 441 KB - jmir_v24i1e28647_app2.pdf]

Multimedia Appendix 3
Changed cut-off points.
[PDF File (Adobe PDF File), 444 KB - jmir_v24i1e28647_app3.pdf]

Multimedia Appendix 4
Well-being scale.
[PNG File , 70 KB - jmir_v24i1e28647_app4.png]

Multimedia Appendix 5
Correlations of phone calls.
[PNG File , 2537 KB - jmir_v24i1e28647_app5.png]

Multimedia Appendix 6
Correlations of other calls.
[PNG File , 1808 KB - jmir_v24i1e28647_app6.png]

Multimedia Appendix 7
Summary of statistics.
[PDF File (Adobe PDF File), 629 KB - jmir_v24i1e28647_app7.pdf]

Multimedia Appendix 8
Boxplots of duration of all calls.
[PNG File , 401 KB - jmir_v24i1e28647_app8.png]

Multimedia Appendix 9
Boxplots of incoming calls.
[PNG File , 346 KB - jmir_v24i1e28647_app9.png]

Multimedia Appendix 10
Boxplots of outgoing calls.
[PNG File , 351 KB - jmir_v24i1e28647_app10.png]

Multimedia Appendix 11
Boxplots of all calls.
[PNG File , 365 KB - jmir_v24i1e28647_app11.png]

Multimedia Appendix 12
Boxplots of duration of incoming calls.
[PNG File , 388 KB - jmir_v24i1e28647_app12.png]

Multimedia Appendix 13
Boxplots of duration of outgoing calls.
[PNG File , 391 KB - jmir_v24i1e28647_app13.png]

Multimedia Appendix 14
Boxplots of fraction of missed calls.
[PNG File , 340 KB - jmir_v24i1e28647_app14.png]

Multimedia Appendix 15
Boxplots of standard deviations of incoming call duration.
[PNG File , 358 KB - jmir_v24i1e28647_app15.png]

Multimedia Appendix 16
Boxplots of standard deviations of outgoing call duration.
[PNG File , 377 KB - jmir_v24i1e28647_app16.png]

Multimedia Appendix 17
Boxplots of fraction of outgoing calls.
[PNG File , 377 KB - jmir_v24i1e28647_app17.png]

Multimedia Appendix 18
Boxplots of number of missed calls.
[PNG File , 331 KB - jmir_v24i1e28647_app18.png]

Multimedia Appendix 19
Boxplots of self-assessment of feeling, sleep time, number, and length of text messages.
[PDF File (Adobe PDF File), 1296 KB - jmir_v24i1e28647_app19.pdf]

Multimedia Appendix 20
Regression coefficients from mixed regression models regarding smartphone-based data and affective states.
[PDF File (Adobe PDF File), 443 KB - jmir_v24i1e28647_app20.pdf]

References

1. Catalá-López F, Gènova-Maleras R, Vieta E, Tabarés-Seisdedos R. The increasing burden of mental and neurological disorders. Eur Neuropsychopharmacol 2013 Nov;23(11):1337-1339. [doi: 10.1016/j.euroneuro.2013.04.001] [Medline: 23643344]
2. Grande I, Berk M, Birmaher B, Vieta E. Bipolar disorder. Lancet 2016 Apr 09;387(10027):1561-1572. [doi: 10.1016/S0140-6736(15)00241-X] [Medline: 26388529]
3. Vázquez GH, Holtzman JN, Lolich M, Ketter TA, Baldessarini RJ. Recurrence rates in bipolar disorder: Systematic comparison of long-term prospective, naturalistic studies versus randomized controlled trials. Eur Neuropsychopharmacol 2015 Oct;25(10):1501-1512. [doi: 10.1016/j.euroneuro.2015.07.013] [Medline: 26238969]
4. Kessing LV, Andersen PK. Predictive effects of previous episodes on the risk of recurrence in depressive and bipolar disorders. Curr Psychiatry Rep 2005 Dec;7(6):413-420. [doi: 10.1007/s11920-005-0061-0] [Medline: 16318818]
5. Park DY, Do D, Chang L, Shah S, Yuen LD, Hooshmand F, et al. Episode accumulation associated with hastened recurrence and delayed recovery in bipolar disorder. J Affect Disord 2018 Feb;227:657-664. [doi: 10.1016/j.jad.2017.11.071] [Medline: 29174739]
6. Vieta E, Berk M, Schulze TG, Carvalho AF, Suppes T, Calabrese JR, et al. Bipolar disorders. Nat Rev Dis Primers 2018 Mar 08;4:18008. [doi: 10.1038/nrdp.2018.8] [Medline: 29516993]
7. Faurholt-Jepsen M, Frost M, Vinberg M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as objective measures of bipolar disorder symptoms. Psychiatry Res 2014 Jun 30;217(1-2):124-127. [doi: 10.1016/j.psychres.2014.03.009] [Medline: 24679993]

8. Faurholt-Jepsen M, Vinberg M, Frost M, Christensen EM, Bardram JE, Kessing LV. Smartphone data as an electronic biomarker of illness activity in bipolar disorder. Bipolar Disord 2015 Nov;17(7):715-728. [doi: 10.1111/bdi.12332] [Medline: 26395972]
9. Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Tröster G, et al. Using smart phone mobility traces for the diagnosis of depressive and manic episodes in bipolar patients. 2014 Presented at: Proceedings of the 5th Augmented Human International Conference; March; Kobe, Japan p. 1-8. [doi: 10.1145/2582051.2582089]
10. Grünerbl A, Muaremi A, Osmani V, Bahle G, Ohler S, Tröster G, et al. Smartphone-based recognition of states and state changes in bipolar disorder patients. IEEE J Biomed Health Inform 2015 Jan;19(1):140-148. [doi: 10.1109/JBHI.2014.2343154] [Medline: 25073181]
11. Faurholt-Jepsen M, Busk J, Frost M, Vinberg M, Christensen EM, Winther O, et al. Voice analysis as an objective state marker in bipolar disorder. Transl Psychiatry 2016 Jul 19;6:e856 [FREE Full text] [doi: 10.1038/tp.2016.123] [Medline: 27434490]
12. Faurholt-Jepsen M, Vinberg M, Frost M, Debel S, Margrethe Christensen E, Bardram JE, et al. Behavioral activities collected through smartphones and the association with illness activity in bipolar disorder. Int J Methods Psychiatr Res 2016 Dec;25(4):309-323 [FREE Full text] [doi: 10.1002/mpr.1502] [Medline: 27038019]
13. Carr O, Saunders KEA, Bilderbeck AC, Tsanas A, Palmius N, Geddes JR, et al. Desynchronization of diurnal rhythms in bipolar disorder and borderline personality disorder. Transl Psychiatry 2018 Apr 12;8(1):79 [FREE Full text] [doi: 10.1038/s41398-018-0125-7] [Medline: 29643339]
14. Monteith S, Glenn T, Geddes J, Bauer M. Big data are coming to psychiatry: a general introduction. Int J Bipolar Disord 2015 Dec;3(1):21 [FREE Full text] [doi: 10.1186/s40345-015-0038-9] [Medline: 26440506]
15. Crisan C. Lack of insight in bipolar disorder: the impact on treatment adherence, adverse clinical outcomes and quality of life. In: Psychotic Disorders-An Update. Budapest: IntechOpen; 2018.
16. Antosik-Wójcińska AZ, Dominiak M, Chojnacka M, Kaczmarek-Majer K, Opara KR, Radziszewska W, Święcicki. Smartphone as a monitoring tool for bipolar disorder: a systematic review including data analysis, machine learning algorithms and predictive modelling. Int J Med Inform 2020 Jun;138:104131. [doi: 10.1016/j.ijmedinf.2020.104131] [Medline: 32305023]
17. Rajagopalan A, Shah P, Zhang MW, Ho RC. Digital Platforms in the Assessment and Monitoring of Patients with Bipolar Disorder. Brain Sci 2017 Nov 12;7(11):150 [FREE Full text] [doi: 10.3390/brainsci7110150] [Medline: 29137156]
18. Torous J, Powell AC. Current research and trends in the use of smartphone applications for mood disorders. Internet Interventions 2015 May;2(2):169-173. [doi: 10.1016/j.invent.2015.03.002]
19. Rucci P, Calugi S, Miniati M, Fagiolini A. A review of self-report and interview-based instruments to assess mania and hypomania symptoms. Giornale Italiano di Psicopatologia. 2013. URL: https://www.researchgate.net/publication/259464128_A_review_of_self-report_and_interview-based_instruments_to_assess_mania_and_hypomania_symptoms [accessed 2021-01-10]
20. Francis A, Gasparo P. Interval between symptom onset and hospitalization in mania. Journal of Affective Disorders 1994 Jul;31(3):179-185. [doi: 10.1016/0165-0327(94)90027-2]
21. Vieta E, Salagre E, Grande I, Carvalho AF, Fernandes BS, Berk M, et al. Early Intervention in Bipolar Disorder. Am J Psychiatry 2018 May 01;175(5):411-426. [doi: 10.1176/appi.ajp.2017.17090972] [Medline: 29361850]
22. Velligan DI, Weiden PJ, Sajatovic M, Scott J, Carpenter D, Ross R, et al. Assessment of adherence problems in patients with serious and persistent mental illness: recommendations from the Expert Consensus Guidelines. J Psychiatr Pract 2010 Jan;16(1):34-45. [doi: 10.1097/01.pra.0000367776.96012.ca] [Medline: 20098229]
23. Daus H, Kislicyn N, Heuer S, Backenstrass M. Disease management apps and technical assistance systems for bipolar disorder: Investigating the patients´ point of view. J Affect Disord 2018 Mar 15;229:351-357. [doi: 10.1016/j.jad.2017.12.059] [Medline: 29331693]
24. Van Til K, McInnis MG, Cochran A. A comparative study of engagement in mobile and wearable health monitoring for bipolar disorder. Bipolar Disord 2020 Mar;22(2):182-190 [FREE Full text] [doi: 10.1111/bdi.12849] [Medline: 31610074]
25. Bardram J, Frost M, Szántó K, Faurholt-Jepsen M, Vinberg M, Kessing L. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. New York, USA: Association for Computing Machinery; 2013 Presented at: SIGCHI Conference on Human Factors in Computing Systems (CHI '13); 2013; Paris, France. [doi: 10.1145/2470654.2481364]
26. Festersen P, Corradini A. Re: Mind: A mobile application for bipolar disorder patients. 2014 Presented at: 4th International Conference on Wireless Mobile Communication and Healthcare - Transforming Healthcare Through Innovations in Mobile and Wireless Technologies (MOBIHEALTH); December 5; Athens, Greece p. 343-346. [doi: 10.4108/icst.mobihealth.2014.257188]
27. Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, Gosling SD. Using Smartphones to Collect Behavioral Data in Psychological Science: Opportunities, Practical Considerations, and Challenges. Perspect Psychol Sci 2016 Nov;11(6):838-854 [FREE Full text] [doi: 10.1177/1745691616650285] [Medline: 27899727]

28. Kessing LV, Hansen MG, Andersen PK, Angst J. The predictive effect of episodes on the risk of recurrence in depressive and bipolar disorders - a life-long perspective. Acta Psychiatr Scand 2004 May;109(5):339-344. [doi: 10.1046/j.1600-0447.2003.00266.x] [Medline: 15049770]
29. Kessing LV, Hansen MG, Andersen PK. Course of illness in depressive and bipolar disorders. Naturalistic study, 1994-1999. Br J Psychiatry 2004 Nov;185:372-377. [doi: 10.1192/bjp.185.5.372] [Medline: 15516544]
30. Karam ZN, Provost EM, Singh S, Montgomery J, Archer C, Harrington G, et al. Ecologically valid long-term mood monitoring of individuals with bipolar disorder using speech. Proc IEEE Int Conf Acoust Speech Signal Process 2014 May;2014:4858-4862 [FREE Full text] [doi: 10.1109/ICASSP.2014.6854525] [Medline: 27630535]
31. Muaremi A, Gravenhorst F, Gruenerbl A, Arnrich B, Troster G. Assessing bipolar episodes using speech cues derived from phone calls. 2014 Presented at: Interntional Symposium on Pervasive Computing Paradigms for Mental Health (mindCare); 2014; Tokyo, Japan p. 103-114. [doi: 10.1007/978-3-319-11564-1_11]
32. Kaminska O, Kaczmarek-Majer K, Opara K, Jakuczun W, Dominiak M, Antosik-Wójcinska A, et al. Self-organizing maps using acoustic features for prediction of state change in bipolar disorder. 2019 Presented at: Proceedings of the 17th Conference on Artificial Intelligence in Medicine, AIME; March; Poznan, Poland. [doi: 10.1007/978-3-030-37446-4_12]
33. Kaminska O, Kaczmarek-Majer K, Hryniewicz O. Acoustic Feature Selection with Fuzzy Clustering, Self Organizing Maps and Psychiatric Assessments. 2020 Presented at: Proceedings of the IPMU 2020: Information Processing and Management of Uncertainty in Knowledge-Based Systems; June 15-19; Lisbon, Portugal p. 342-355. [doi: 10.1007/978-3-030-50146-4_26]
34. Luke SG. Evaluating significance in linear mixed-effects models in R. Behav Res Methods 2017 Aug;49(4):1494-1502. [doi: 10.3758/s13428-016-0809-y] [Medline: 27620283]
35. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. J Stat Soft 2015;67(1):1-48. [doi: 10.18637/jss.v067.i01]
36. Beiwinkel T, Kindermann S, Maier A, Kerl C, Moock J, Barbian G, et al. Using Smartphones to Monitor Bipolar Disorder Symptoms: A Pilot Study. JMIR Ment Health 2016 Jan 06;3(1):e2 [FREE Full text] [doi: 10.2196/mental.4560] [Medline: 26740354]
37. Johnson SL, Morriss R, Scott J, Paykel E, Kinderman P, Kolamunnage-Dona R, et al. Depressive and manic symptoms are not opposite poles in bipolar disorder. Acta Psychiatr Scand 2011 Mar;123(3):206-210 [FREE Full text] [doi: 10.1111/j.1600-0447.2010.01602.x] [Medline: 20825373]
38. Depp CA, Kim DH, de Dios LV, Wang V, Ceglowski J. A Pilot Study of Mood Ratings Captured by Mobile Phone Versus Paper-and-Pencil Mood Charts in Bipolar Disorder. J Dual Diagn 2012 Jan 01;8(4):326-332 [FREE Full text] [doi: 10.1080/15504263.2012.723318] [Medline: 23646035]
39. Faurholt-Jepsen M, Torri E, Cobo J, Yazdanyar D, Palao D, Cardoner N, et al. Smartphone-based self-monitoring in bipolar disorder: evaluation of usability and feasibility of two systems. Int J Bipolar Disord 2019 Jan 04;7(1):1 [FREE Full text] [doi: 10.1186/s40345-018-0134-8] [Medline: 30610400]
40. Busk J, Faurholt-Jepsen M, Frost M, Bardram JE, Kessing LV, Winther O. Daily estimates of clinical severity of symptoms in bipolar disorder from smartphone-based self-assessments. Transl Psychiatry 2020 Jun 18;10(1):194 [FREE Full text] [doi: 10.1038/s41398-020-00867-6] [Medline: 32555144]
41. Faurholt-Jepsen M, Vinberg M, Christensen EM, Frost M, Bardram J, Kessing LV. Daily electronic self-monitoring of subjective and objective symptoms in bipolar disorder--the MONARCA trial protocol (MONitoring, treAtment and pRediCtion of bipolAr disorder episodes): a randomised controlled single-blind trial. BMJ Open 2013;3(7):e003353 [FREE Full text] [doi: 10.1136/bmjopen-2013-003353] [Medline: 23883891]
42. Hidalgo-Mazzei D, Mateu A, Reinares M, Murru A, Del Mar Bonnín C, Varo C, et al. Psychoeducation in bipolar disorder with a SIMPLe smartphone application: Feasibility, acceptability and satisfaction. J Affect Disord 2016 Aug;200:58-66. [doi: 10.1016/j.jad.2016.04.042] [Medline: 27128358]
43. Hidalgo-Mazzei D, Reinares M, Mateu A, Juruena MF, Young AH, Pérez-Sola V, et al. Is a SIMPLe smartphone application capable of improving biological rhythms in bipolar disorder? J Affect Disord 2017 Dec 01;223:10-16. [doi: 10.1016/j.jad.2017.07.028] [Medline: 28711743]
44. Hidalgo-Mazzei D, Reinares M, Mateu A, Nikolova VL, Bonnín CDM, Samalin L, et al. OpenSIMPLe: A real-world implementation feasibility study of a smartphone-based psychoeducation programme for bipolar disorder. J Affect Disord 2018 Dec 01;241:436-445. [doi: 10.1016/j.jad.2018.08.048] [Medline: 30145515]
45. Chan EC, Sun Y, Aitchison KJ, Sivapalan S. Mobile App-Based Self-Report Questionnaires for the Assessment and Monitoring of Bipolar Disorder: Systematic Review. JMIR Form Res 2021 Jan 08;5(1):e13770 [FREE Full text] [doi: 10.2196/13770] [Medline: 33416510]
46. Bonnín CDM, Solé B, Reinares M, García-Estela A, Samalin L, Martínez-Arán A, et al. Does cognitive impairment in bipolar disorder impact on a SIMPLe app use? J Affect Disord 2021 Mar 01;282:488-494. [doi: 10.1016/j.jad.2020.12.168] [Medline: 33422826]
47. Gideon J, Provost EM, McInnis M. Mood state prediction from speech of varying acoustic quality for individuals with bipolar disorder. Proc IEEE Int Conf Acoust Speech Signal Process 2016 Mar;2016:2359-2363 [FREE Full text] [doi: 10.1109/ICASSP.2016.7472099] [Medline: 27570493]

48. Cummins N, Scherer S, Krajewski J, Schnieder S, Epps J, Quatieri TF. A review of depression and suicide risk assessment using speech analysis. Speech Communication 2015 Jul;71:10-49. [doi: 10.1016/j.specom.2015.03.004]
49. Fukazawa Y, Yamamoto N, Hamatani T, Ochiai K, Uchiyama A, Ohta K. Smartphone-based Mental State Estimation: A Survey from a Machine Learning Perspective. Journal of Information Processing 2020;28:16-30. [doi: 10.2197/ipsjjip.28.16]
50. Vazquez-Montes MDLA, Stevens R, Perera R, Saunders K, Geddes JR. Control charts for monitoring mood stability as a predictor of severe episodes in patients with bipolar disorder. Int J Bipolar Disord 2018 Apr 04;6(1):7 [FREE Full text] [doi: 10.1186/s40345-017-0116-2] [Medline: 29616434]

Abbreviations

BD: bipolar disorder
HDRS: Hamilton Depression Rating Scale
ICD-10: International Classification of Diseases, tenth revision
YMRS: Young Mania Rating Scale

Edited by R Kukafka; submitted 14.03.21; peer-reviewed by D Hidalgo-Mazzei, E Chan; comments to author 21.04.21; revised version received 15.06.21; accepted 15.11.21; published 19.01.22.

Please cite as:
Dominiak M, Kaczmarek-Majer K, Antosik-Wójcińska AZ, Opara KR, Olwert A, Radziszewska W, Hryniewicz O, Święcicki Ł, Wojnar M, Mierzejewski P
Behavioral and Self-reported Data Collected From Smartphones for the Assessment of Depressive and Manic Symptoms in Patients With Bipolar Disorder: Prospective Observational Study
J Med Internet Res 2022;24(1):e28647
URL: https://www.jmir.org/2022/1/e28647
doi:10.2196/28647
PMID:34874015

Original Paper

Encouraging Behavior Changes and Preventing Cardiovascular Diseases Using the Prevent Connect Mobile Health App: Conception and Evaluation of App Quality

Dahbia Agher[1,2], PhD; Karima Sedki[1], PhD; Sylvie Despres[1], PhD; Jean-Pierre Albinet[2]; Marie-Christine Jaulent[1], PhD; Rosy Tsopra[3,4,5], MD, PhD

[1]Inserm, University Sorbonne Paris Nord, Sorbonne University, Laboratory of Medical Informatics and Knowledge Engineering in e-Health, LIMICS, Paris, France

[2]BeWellConnect Research and Development, Visiomed Group, Puteaux, France

[3]Inserm, Université de Paris, Sorbonne Université, Centre de Recherche des Cordeliers, F-75006, Paris, France

[4]HEKA, Inria, Paris, France

[5]Department of Medical Informatics, Hôpital Européen Georges-Pompidou, Assistance Publique - Hôpitaux de Paris, Paris, France

Corresponding Author:
Dahbia Agher, PhD
Inserm, University Sorbonne Paris Nord, Sorbonne University
Laboratory of Medical Informatics and Knowledge Engineering in e-Health, LIMICS
15 Rue de l'école de Médecine
Paris, 75006
France
Phone: 33 0619260049
Email: agherdahbia2@gmail.com

Abstract

Background: Cardiovascular diseases are a major cause of death worldwide. Mobile health apps could help in preventing cardiovascular diseases by improving modifiable risk factors such as eating habits, physical activity levels, and alcohol or tobacco consumption.

Objective: The aim of this study was to design a mobile health app, Prevent Connect, and to assess its quality for (1) assessing patient behavior for 4 cardiovascular risk factors (unhealthy eating, sedentary lifestyle, alcohol, and tobacco consumption) and (2) suggesting personalized recommendations and mobile health interventions for risky behaviors.

Methods: The knowledge base of the app is based on French national recommendations for healthy eating, physical activity, and limiting alcohol and tobacco consumption. It contains a list of patient behaviors and related personalized recommendations and digital health interventions. The interface was designed according to usability principles. Its quality was assessed by a panel of 52 users in a 5-step process: completion of the demographic form, visualization of a short presentation of the app, testing of the app, completion of the user version of the Mobile App Rating Scale (uMARS), and an open group discussion.

Results: This app assesses patient behaviors through specific questionnaires about 4 risk factors (unhealthy eating, sedentary lifestyle, alcohol, and tobacco consumption) and suggests personalized recommendations and digital health interventions for improving behavior. The app was deemed to be of good quality, with a mean uMARS quality score of 4 on a 5-point Likert scale. The functionality and information content of the app were particularly appreciated, with a mean uMARS score above 4. Almost all the study participants appreciated the navigation system and found the app easy to use. More than three-quarters of the study participants found the app content relevant, concise, and comprehensive. The aesthetics and the engagement of the app were also appreciated (uMARS score, 3.7). Overall, 80% (42/52) of the study participants declared that the app helped them to become aware of the importance of addressing health behavior, and 65% (34/52) said that the app helped motivate them to change lifestyle habits.

Conclusions: The app assessed the risky behaviors of the patients and delivered personalized recommendations and digital health interventions for multiple risk factors. The quality of the app was considered to be good, but the impact of the app on behavior changes is yet to be demonstrated and will be assessed in further studies.

(J Med Internet Res 2022;24(1):e25384) doi:10.2196/25384

KEYWORDS

digital health; mHealth, mobile application; IT; technology; prevention; cardiovascular risk factor; behavior change; primary care

Introduction

Cardiovascular diseases (CVDs) are a major cause of death worldwide [1] for which several causal risk factors have been identified [2]. One of the keys to prevent CVD is reducing the impact of modifiable risk factors such as unhealthy eating, physical inactivity, alcohol consumption, and smoking [3-7]. This can be achieved by providing advice and encouraging changes in patient behavior, particularly as concerns these risk factors. Mobile health (mHealth) interventions [8-10] can be useful for improving lifestyle behaviors relating to CVDs. mHealth technology is defined as mobile devices such as mobile phones, patient monitoring devices, personal digital assistants, or other wireless devices intended to be worn, carried, or accessed by patients or health care providers to monitor health status or improve health outcomes [3,4].

mHealth interventions may take various forms, including SMS text messages, interactive voice responses, surveys, focus groups, smartphone health apps, and solutions combining several aspects of connected health. Many of these interventions have been shown to have a significant impact on the management and improvement of modifiable risk factors in CVD prevention [5-7,11-14]. For example, many studies [14] have reported promising results for mHealth interventions in the improvement of patient outcomes such as body measurements (eg, weight, waist circumference), metabolic and physiological measurements (eg, blood pressure, glucose levels), adherence to and safe use of medication, physical activity performance, meal management, and awareness of health conditions and treatment options.

Despite the emergence of mHealth and research into lifestyle behavior changes, most of the interventions developed to date target one single risk factor for CVD prevention. Positive effects of mHealth interventions have been demonstrated for single risk factors for CVD, for example, Text2Quit [15] for smoking cessation, ASA24 [16] for recording food intake, and mDiet [17] for weight loss and for alcohol intake [18]. Some studies [14] have also examined the impact of mHealth interventions on multiple risk factors, but these studies have reported only limited success. These studies were limited by a lack of personalization for either delivering recommendations or for selecting the most appropriate form of intervention (SMS, email, etc) as a function of patient profile. Furthermore, many of the apps are at risk of becoming rapidly obsolete owing to the fast pace at which technologies are progressing, and new technological innovations must therefore be considered. For example, the latest mobile technologies can connect and interact with each other, update, and track personal health data in real time, and send alerts to users. In addition, the number of users with a modern smartphone has considerably increased in recent years (from 35% in 2011 to 81% in 2019 [9]). Likewise, most health apps have encountered serious usability problems [19] or have not undergone usability assessment [20]. Usability affects the efficiency and efficacy of the app (eg, time to complete tasks, errors) [21] and must be considered to increase the chance of the app being successfully adopted by patients.

Here, we designed an mHealth app, Prevent Connect, for (1) assessing patient behaviors for 4 important risk factors (unhealthy eating, sedentary lifestyle, alcohol, and tobacco consumption) and (2) suggesting personalized recommendations and mHealth interventions for risky behaviors. We aimed to design an app with sufficiently high usability to favor its adoption by patients. We describe the Prevent Connect app and the evaluation of its quality and usability.

Methods

Design of Prevent Connect

Knowledge Base

The knowledge base of the mobile app was based on French national recommendations for healthy eating, physical activity levels, and limiting alcohol and tobacco consumption [22-25]. We analyzed these recommendations and extracted the necessary variables for the assessments of each behavior (eg, for healthy eating, we extracted 23 variables, each with 7 possible answers). We then combined the variables and their responses to establish a list of patient behaviors. For each patient behavior, we calculated a qualitative score extending from “very good behavior” to “very bad behavior” (based on guidelines). We then associated personalized recommended actions with these behavior scores (eg, “do the equivalent of 30 minutes of physical activity per day”) and digital health interventions (eg, “activity tracker for self-monitoring of daily activity”). The identified patient behaviors and their associated qualitative scores for risky behavior, personalized recommendations, and digital health interventions were then implemented within the app.

Interface

The mobile app was ergonomically designed according to usability principles [26,27], and the graphical interface was implemented with Gluon mobile technology [28]. The navigation system was based on a 3-step process: (1) assessment of each risk factor with specific questionnaires, (2) automatic assessment, by the app, of risky behavior, and visualization of recommended and personalized actions, and (3) suggestions for digital health interventions. The interface was designed according to the following usability principles [26,27]: simplicity, naturalness, and effective use of language by using concise, appropriate, and understandable language; consistency, minimizing cognitive load, and efficient interactions by dividing the navigation into a 3-step process and adding a clear taskbar and meaningful colors; and effective information presentation, by limiting the amount of text and using comprehensible icons (eg, a wineglass icon was used to represent alcohol consumption). The consideration of usability principles in app design can help reduce the user’s cognitive workload and increase confidence in the app [29].

Evaluation of Prevent Connect

General Study Design and Ethics Approval

We carried out an evaluation of the quality of the app by asking patients identified as potential future users to test the app. We organized 20 online sessions, each with 1-5 users. Each session began with a demonstration of the app and then testing of the app by the study participants. The opinions of the study participants were then collected in an electronic form and through an open discussion. The study protocol was validated by the appropriate Inserm ethics committee (CD/EB 20-023, 20-660, IRB000388, IORG0003254, FWA0005831).

Recruitment

Study participants were recruited online via a dedicated website and by word-of-mouth from April 15 to June 15, 2020. For participation, subjects had to be older than 18 years and not treated for CVD (eg, stroke, coronary artery disease). They also had to sign an online consent form and agree to participate without compensation.

Study Design and Statistical Analysis

The evaluation was performed online and lasted from 30 minutes to 1 hour, depending on the session. It proceeded via 5 successive steps:

1. In step 1, study participants completed an online form to provide sociodemographic information. Questions were adapted so as to guarantee anonymity, and no personal information was recorded (eg, participants were asked to indicate their age group rather than an exact age to limit the risks of reidentification).
2. In step 2, study participants watched a short demonstration of the app.
3. In step 3, study participants used and tested the app on their own. They first completed the questionnaires for behavior assessment and visualized the recommendations delivered by the app.
4. In step 4, study participants completed the user version of the Mobile App Rating Scale (uMARS) evaluation form online, together with 2 additional questions about satisfaction and possible areas of improvement. The uMARS form contains 26 items for assessing app quality [30]: 5 items for app engagement, 4 items for app functionality, 3 items for aesthetics, 4 items for app information, 4 items for the subjective quality of the app, and 6 items for the perceived impact of the app. The global quality of the app was then assessed by calculating the mean score for the uMARS items corresponding to the categories engagement, functionality, aesthetics, and information, according to the following formula [31]: mean score = (engagement score + functionality score + aesthetics score + information score)/4, with a maximum mean score of 5.
5. In step 5, study participants were invited to discuss the app freely in groups. They were allowed to give their opinions freely and to discuss the app together.

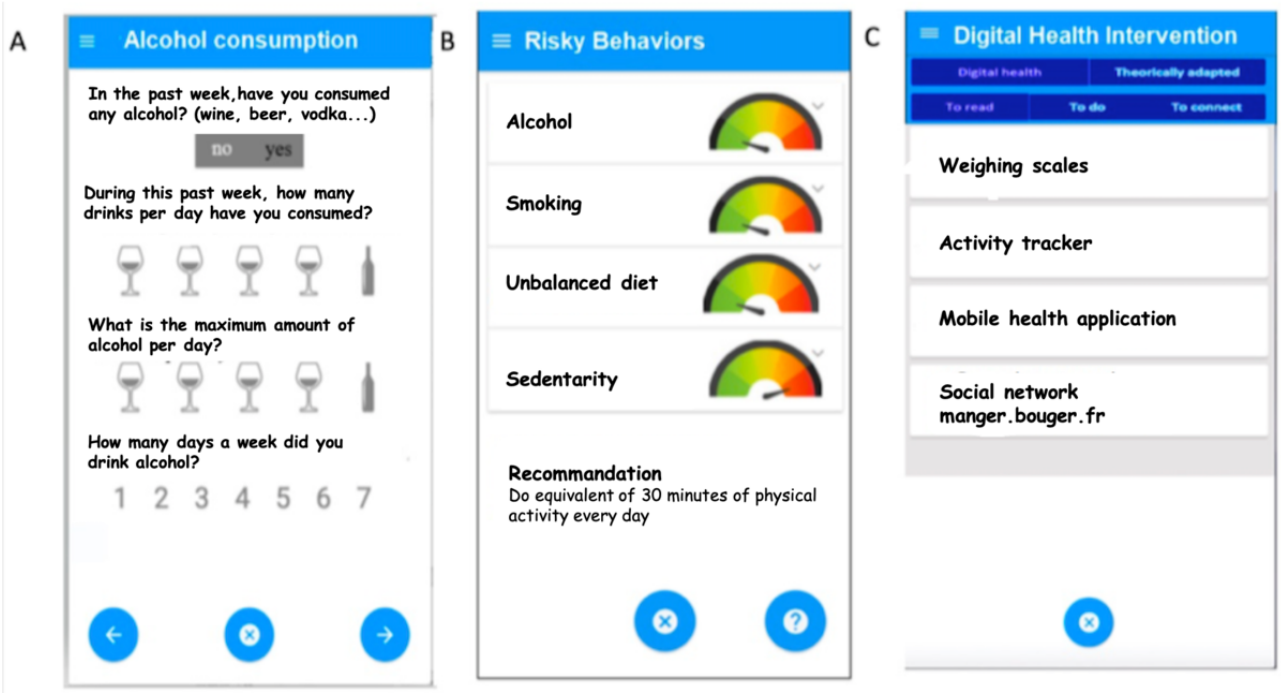
For sociodemographic data and for each item of the uMARS scale, we calculated the percentage of each response. For the free discussion, data were analyzed, broken up, and combined into similar themes. Each theme was mapped to one of the categories of the uMARS scale (eg, engagement, functionality).

Results

The Prevent Connect App

Use of the app begins with the login interface (Figure 1). Once the users have logged in, they visualize the home interface, which includes several tabs (eg, welcome first time, behavior questionnaire). During the first connection, patients are asked a number of questions about themselves (eg, gender, age, family history, social context) and behaviors concerning eating habits, physical activity, and tobacco and alcohol consumption. Once these forms have been completed, the app provides a recap of the behavior assessment for each risk factor in the form of a gauge with an arrow indicating the position of the user on a color scale extending from green for “very good behavior” to red for “very bad behavior” (graduated scale depending on the qualitative score calculated for risky behavior). Personalized and recommended actions regarding behavior changes are also displayed at the bottom of the interface. Users then go on to the next screen, which displays the personalized digital health interventions that would be helpful in encouraging behavior changes. A help icon is also provided to allow users to obtain additional explanations or information on request.

Figure 1. Illustration of the "Prevent Connect" app. On the first interface, users complete questionnaires to assess their behavior for each risk factor (eg, in area A, the alcohol consumption questionnaire is shown). The app automatically identifies risky behaviors and the users can consult this result and the recommended actions on the second interface (eg, in area B, the user is sedentary, and a personalized recommendation is displayed). On the last interface, users can consult a list of targeted digital interventions that are personalized according to their risky behaviors (eg, in area C, an activity tracker or a weighing scale is suggested because the patient is sedentary). Note: manger.bouger.fr is a well-known social network in France that provides additional tips and advice regarding physical activities and healthy eating.



Evaluation of the Quality of the Prevent Connect App

Description of the Panel of Study Participants

In total, 52 individuals tested and evaluated the app: 28 women and 24 men. The study participants were aged between 18 and 69 years and had diverse professional profiles. Most participants were familiar with digital devices and used a smartphone, computer tablet, or laptop computer at least once per week. Half of the participants declared that they had already used a connected object, mostly with an app, but only 13% (7/52) declared that they had already assessed their cardiovascular risk in this way (Table 1).

Table 1. Sociodemographic characteristics of the study participants (N=52).

Characteristics	Values, n (%)
Sex	
Female	28 (54)
Male	24 (46)
Age (years)	
18-24	8 (15)
25-34	23 (44)
35-49	16 (31)
50-69	5 (10)
Socioprofessional category	
Student	11 (21)
Executive	24 (46)
Employee	2 (4)
Liberal profession	9 (17)
Unemployed	4 (8)
Retired	2 (4)
Use of smartphone, laptop computer, or computer tablet at least once per week	
Smartphone only	4 (8)
Laptop computer only	2 (4)
Computer tablet only	0 (0)
Smartphone and laptop computer	29 (56)
Smartphone and computer tablet	2 (4)
Smartphone, computer tablet, and laptop computer	15 (29)
Use of connected objects (eg, activity tracker)	
Yes, at least once per day	10 (19)
Yes, at least once per week	7 (13)
Yes, at least once per month	4 (8)
Yes, less than once per month	6 (12)
No use	25 (48)
Mode of connected object use	
Exclusively with the help of a health professional	1 (2)
Exclusively with an app	23 (44)
By turning the Wi-Fi off	1 (2)
With a health professional and app	2 (4)
No use	25 (48)
Evaluation of cardiovascular risk (present or past)	
Yes, less than once per month	7 (13)
No	45 (87)
Mode of cardiovascular risk evaluation	
Health professional	3 (6)
Apps	2 (4)
Website	1 (2)
Score	1 (2)

Characteristics	Values, n (%)
No item	45 (86)

App Quality

The mean total score for app quality was 4 on the uMARS scale (maximum mean score of 5, Figure 2).

For *engagement*, the mean uMARS score was 3.7 (SD 0.7). More than three-quarters of the study participants found the app interesting and appropriate for the target audience, and more than half found it fun and entertaining (Table 2, comment 1). One study participant said that the app was appropriate and accessible to everyone, including older adults who were less used to apps (Table 2, comment 2). However, fewer than half of the study participants found the app sufficiently interactive and customizable. Some study participants suggested making the app more interactive by adding a chatbot and the possibility of speaking to an expert, together with additional reminders and alerts concerning the recommended behavior changes. A few study participants also suggested delivering a daily preventive message to help educate users (Table 2, comment 3).

For *functionality*, the mean uMARS score was 4.4 (SD 0.6). Almost all the study participants appreciated the navigation system and found the app easy to use. More than 80% of the participants also appreciated the performance and design of the app (Table 2, comment 4). Study participants said that they found the app intuitive and very easy to use, although 1 person suggested adding a tutorial (Table 2, comment 5). None of the study participants declared having experienced technical problems during the evaluation.

Figure 2. Distribution of users' responses to the user version of the Mobile App Rating Scale (N=52, except for the item "credibility of source," for which only 51 answers were available, and for the item "visual information," for which only 48 answers were available).

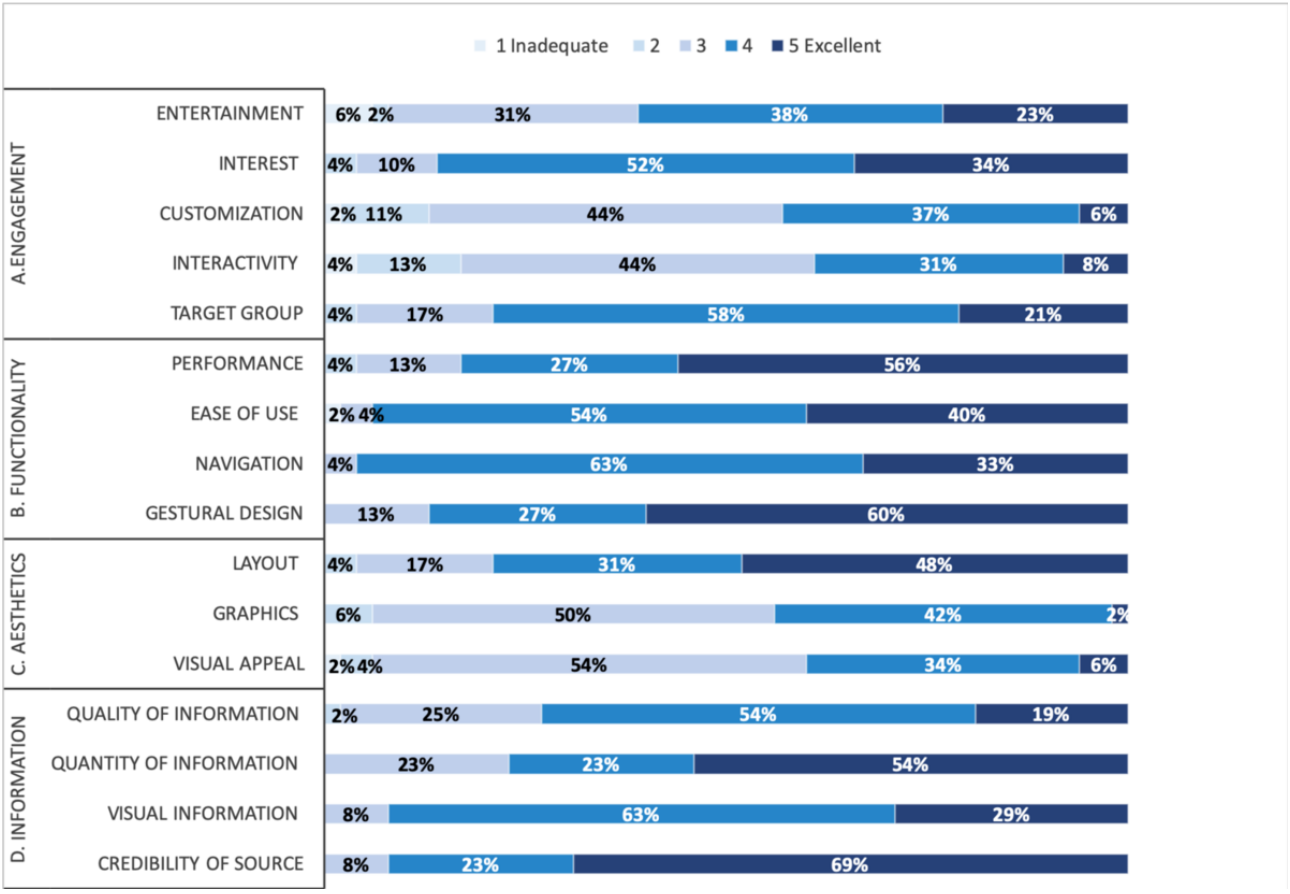


Table 2. Examples of feedback from study participants.

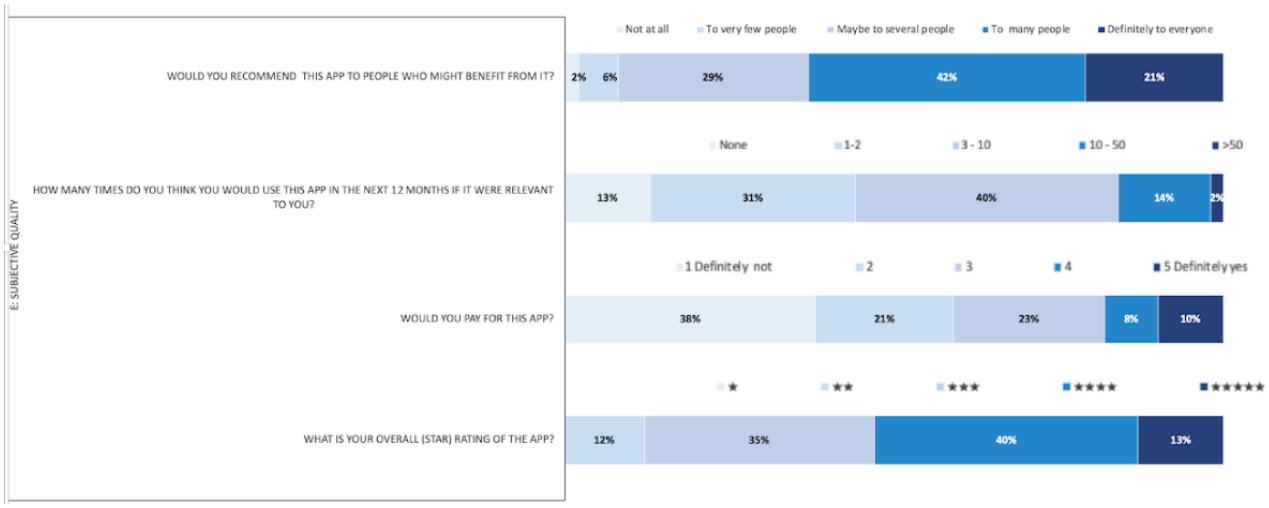
Section, subsection	Extract of comments
Engagement	
Entertainment and interest	…*The mobile app is fun and enables us to assess our behaviors.* [Comment 1]
Customization and interactivity	…*The app could be more interactive and customizable with a chatbot, the possibility to talk with an expert, but also with reminders and alerts about the recommended behavior changes.* [Comment 2]
Target group	…*The app is appropriate and accessible to everybody, even for older adults less accustomed to apps.* [Comment 3]
Functionality	
Performance	…*I didn't have any technical problems when I used the app.* [Comment 4]
Ease of use and navigation	…*Very practical and easy to use. The app can be used to make a quick assessment, but also encourages behavior changes, such as doing more sport or adapting food quantity.* [Comment 5]
Aesthetics	
Layout	…*The aesthetics could be improved by adding 3D pictures, video animations, summaries, pictograms (eg, arrows), and emojis.* [Comment 6]
Graphics, visual	…*Colors and graphics are consistent with a health app. Even if the blue color is currently used for health apps, I would prefer green.* [Comment 7]
Information	
Quality and quantity	…*The app content is of good quality (…). The content is right relevant, concise, interesting, clear and detailed.* [Comment 8]
Visual	…*Regarding app content, you could add shocking images and information like those displayed on cigarette packets or during the campaign for preventing traffic accidents.* [Comment 9]
Credibility of source	…*This app inspires more confidence than apps available on app stores.* [Comment 10]
Subjective	
Quality	…*Yes, this is a good app. I would give it a score of 4 out of 5, and I will use it to check my results and to see if I have achieved the target behaviors.* [Comment 11]
Perceived impact	
Awareness and knowledge	…*This is a very good app for becoming aware of your own behavior and changing your lifestyle, and thus for preventing health problems.* [Comment 12]
Seeking help and behavioral changes	…*Used alone no, but associated with follow-up by a health professional, why not?* [Comment 13]

For *aesthetics*, the mean uMARS score was 3.7 (SD 0.6). More than three-quarters of the study participants found the layout appropriate. However, less than half felt that the graphics were of sufficiently high quality and visually appealing. Some study participants found the design and colors appropriate for a health app, whereas others felt that the colors were insufficiently attractive (Table 2, comment 7). Some study participants suggested improving the aesthetics of the app by adding more 3D pictures, video animations, summaries, pictograms, (eg, arrows), and emojis (Table 2, comment 6).

For *information*, the mean uMARS score was 4.2 (SD 0.5). More than three-quarters of the study participants found the app content appropriate, relevant, concise, and comprehensive (Table 2, comment 8). Study participants said they found the content interesting, clear, and detailed. Almost all the study participants also appreciated the visual elements used to display information and had confidence in the credibility of the source used to build the app content (Table 2, comment 9). One study participant said that this app inspired more confidence than those available on app stores (Table 2, comment 10). Some study participants suggested improving the questionnaires used to assess cardiovascular risk by adding more accurate questions about family history, tobacco and alcohol use, and physical activities, for example, or by adding questions about drugs and making the eating habit forms easier to complete. Other study participants suggested adding more information about dairy products, sports activities, and recipes. One study participant suggested adding explanations about the workings of the human body and physiology (eg, heart physiology).

For *subjective opinion* about app quality, 92% (48/52) of the study participants said they would recommend the app to several people, 56% (29/52) of the study participants also declared that they would use the app more than 3-10 times in the next 12 months, and 59% (31/52) said that they would not be willing to pay for it. Finally, 88% (46/52) of the participants gave more than 3 stars for the app when asked to rate it (Table 2, comment 11) (Figure 3).

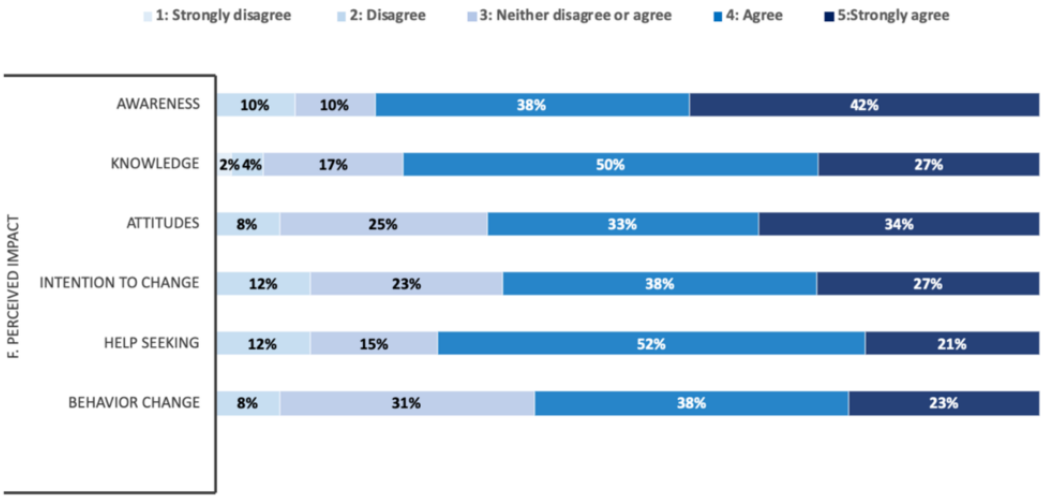
Figure 3. Distribution of users' responses to the user version of the Mobile App Rating Scale for subjective quality (N=52).



Perceived Impact, Global Satisfaction, and Areas of Improvement

For *perceived impact* (Figure 4), 80% (42/52) of the study participants declared that the app had made them more aware of the importance of addressing health behavior, and 77% (40/52) said that it had helped to improve their awareness of their behavior. Some study participants explained that the automatic assessment of behaviors had improved their understanding of their lifestyles and attitudes. One study participant also said that this awareness could encourage changes in behavior to prevent health disorders (Table 2, comment 12). Overall, 67% (35/52) of the participants said that the app had changed their attitudes toward improving health behavior, and 65% (34/52) said that they were more motivated to change their habits by playing more sports or eating more healthily, for example. One study participant said that the app could help him to achieve his personal goals in terms of behavior changes. Furthermore, 61% (32/52) of the participants thought that the app would lead to improvements in health behavior, and 73% (38/52) thought that the app would encourage people to seek further help if required (Table 2, comment 13).

Figure 4. Distribution of users' responses to the user version of the Mobile App Rating Scale for perceived impact (N=52).



For *global satisfaction*, 55% (29/52) of the study participants were fully satisfied with the app, and only 12% (6/52) were not. The others were neutral. Almost all the study participants found that the recommendations delivered were clear.

For *areas of improvement*, some study participants suggested extending the app to cardiovascular follow-up. They suggested adding new functionalities for tracing behavioral changes such as functions for recording meals or an activity tracker for recording physical activity. They also suggested displaying the data collected through their other connected objects (eg, sleep data collected with connected watch) within the app. One participant also suggested adding a virtual avatar to assess progress in behavioral changes over time.

Discussion

Principal Results

We designed Prevent Connect, an app based on national recommendations, to encourage behavioral change for CVD prevention. This app assesses 4 risk factors (unhealthy eating, sedentary lifestyle, alcohol, and tobacco consumption) and delivers personalized recommendations and digital health interventions to encourage changes in behavior. The quality

and usability of the app were assessed by 52 potential future users. They considered the app to be of good quality, with a mean uMARS quality score of 4 on a 5-point Likert scale. The functionality and information content of the app were particularly well appreciated, with uMARS scores above 4. The aesthetics and the engagement of the app were also appreciated (uMARS score 3.7), but further improvements are required.

Strengths and Limitations

This study has several strengths. First, the app is not limited to a single risk factor, but instead covers multiple risk factors. It also delivers recommendations and digital health interventions personalized according to the profile of the patient. Furthermore, the app is available for use on mobile phones, which should facilitate its adoption by patients [32,33]. Second, we considered usability principles [26,27] such as simplicity, naturalness, effective use of language, consistency, minimizing cognitive load, efficient interactions, and effective information presentation to improve the design of the app. It has been shown that the consideration of usability principles can considerably reduce user cognitive workload and increase confidence in technology [29], thereby increasing the likelihood of successful adoption of the app [34,35]. Third, we assessed the quality and usability of the app with a diverse panel of 52 potential users. We used the uMARS scale, which has been validated [31] for the assessment of various aspects of quality (engagement, functionality, aesthetics, and information), and we organized a free discussion with participants. The assessment of both quality and usability is a crucial step in the software lifecycle. Most health apps encounter serious usability problems [19], which may have a negative impact on app efficiency and efficacy (eg, time to complete the tasks, content errors) [21], thereby decreasing the likelihood of app adoption. Usability testing can identify any serious issues early in app development, making it possible to improve the app.

Our study also has several weaknesses. It focuses on primary rather than secondary prevention as it is vital to trigger behavioral changes early before the disease appears to prevent CVD [36]. However, we plan, in the future, to extend the scope of the app to secondary prevention. Another limitation is that we assessed only the quality of the app—not its impact on changes in behavior. Changing behavior is a complex process [37], and even if this app makes patients more aware of their risky behaviors and improves their understanding of healthy behavior, we cannot guarantee that this will be sufficient to have a positive impact on behavior. We plan to set up a randomized controlled trial to assess the impact of this app on changes in behavior and CVD prevention.

Comparison With Other Studies

We propose here interventions for changing behavior at the individual level. Other interventions such as mass media campaigns have been proposed to induce changes in behavior at the population level [38]. They have the advantage of reaching a larger population through the dissemination of prevention messages via the mass media such as television, radio, or newspapers [38]. They also have the potential to produce positive behavior changes and to prevent negative changes [38]. However, they are not personalized to individuals and usually expose the population to the negative consequences of risky behavior, such as the arterial damage caused by smoking (Australian campaign “Every Cigarette is Doing You Damage” [39]). Traditional interventions are designed to evoke fear [32], but it is also important to deliver more positive messages [40]. By doing this, our app may help individuals to change their behavior with greater motivation and enthusiasm, favoring the persistence of behavioral changes over time [41]. Various apps have been developed for cardiovascular prevention at the individual level. However, these apps have several limitations. First, most are limited to a single risk factor. For example, Text2Quit [15] is an interactive mHealth program that sends text messages to offer advice, support, and reminders about quitting smoking, and ASA24 [16] is an automated self-administered recall system for collecting information about food intake. Our app has the advantage of covering multiple risk factors (unhealthy eating habits, sedentary lifestyle, alcohol, and tobacco consumption). Second, most of these apps lack personalization, particularly for the recommendation of digital health interventions (SMS, email, use of connected weighing scales, etc). Personalization is important for ensuring the successful adoption of the recommendations delivered [40]. For example, it has been shown [42] that SMS-based interventions should not be recommended to individuals not familiar with the use of such messages. Costly digital health interventions should not be recommended to patients with financial difficulty, and mobile interactive voice responses should not be recommended if the subject is deaf or unable to respond to questions via a touch-tone phone [42]. Our app has the advantage of delivering free personalized digital health interventions adapted to patient profiles, as defined in the French national health recommendations. In the future, we plan to increase the degree of personalization by considering additional conditions with a potential impact on patient adherence to recommendations (eg, living conditions such as work or family constraints). The consideration of such conditions will require the development of new guidelines, with the help of a multidisciplinary group of experts.

Areas for Improvement

In the future, we aim to incorporate new functionalities, suggested by the participants, into the app. First, some study participants suggested adding functions for tracing behavioral changes. These changes could be monitored with connected objects such as t-shirts, bracelet watches, and smart socks [43]. Connected objects are increasingly being used by individuals for the instantaneous tracking of physiological activities such as physical activity, sleeping, or healthy eating [44]. Such monitoring could be very useful for patient follow-up and for adapting recommendations over time. Furthermore, the tracking achieved with these connected objects can be very effective for encouraging people to change their behavior [45,46]. For these reasons, we plan to connect our app to such tracking devices. Second, some study participants suggested delivering daily messages to support behavioral changes, whereas others suggested adding the possibility to talk with an expert or coach. Alternatively, a conversational agent (or chatbot), a computer program designed to simulate human text or verbal conversations [47,48], could be added. Chatbots have already

been used as an alternative to face-to-face counseling for assisting clinicians, supporting patients trying to change their behavior and for monitoring health conditions [49]. A recent literature review [50] showed that interventions including conversational agents for coaching people in a healthy lifestyle could be more engaging, although data concerning their efficacy remain inconclusive. We will incorporate a conversational agent into our app in the future to increase the chances of behavioral changes being successfully adopted over time.

Conclusions

We developed Prevent Connect, an app based on national recommendations, to encourage behavioral changes for the prevention of CVDs. The app allows users to self-assess the 4 most important risk factors (unhealthy eating habits, sedentary lifestyle, alcohol, and tobacco consumption) and delivers personalized recommendations and digital health interventions to help individuals improve their behavior. The app was considered to be of good quality by a panel of potential future users, but further improvements are required. The impact of the app on behavioral changes remains to be demonstrated and will be assessed in future studies.

Acknowledgments

We thank Dr François Teboul for advice, Akram Redjdal for technical support, and the study participants for their time. This work was supported by BeWellConnect, the Visiomed Group, and the French National Association for Technical Research.

Authors' Contributions

DA, KS, SD, MCJ, and RT made substantial contributions to the design of the study. DA and RT contributed to the acquisition, analysis, and interpretation of the data. All the authors validated the results of the study. DA, KS, and RT drafted the article. All authors critically reviewed the article and approved the final version.

Conflicts of Interest

KS, SD, MCJ, and RT have no conflicts of interest relating to this research to declare. DA and JPA are employed by BeWellConnect.

References

1. Global status report on noncommunicable diseases 2010. World Health Organization. 2012 Jun 16. URL: https://www.who.int/nmh/publications/ncd_report_full_en.pdf [accessed 2022-01-04]
2. Mahmood SS, Levy D, Vasan RS, Wang TJ. The Framingham Heart Study and the epidemiology of cardiovascular disease: a historical perspective. Lancet 2014 Mar 15;383(9921):999-1008 [FREE Full text] [doi: 10.1016/S0140-6736(13)61752-3] [Medline: 24084292]
3. Kumar S, Nilsen WJ, Abernethy A, Atienza A, Patrick K, Pavel M, et al. Mobile health technology evaluation: the mHealth evidence workshop. Am J Prev Med 2013 Aug;45(2):228-236 [FREE Full text] [doi: 10.1016/j.amepre.2013.03.017] [Medline: 23867031]
4. de Vries H, Brug J. Computer-tailored interventions motivating people to adopt health promoting behaviours: introduction to a new approach. Patient Educ Couns 1999 Feb;36(2):99-105. [doi: 10.1016/s0738-3991(98)00127-x] [Medline: 10223015]
5. Foster C, Richards J, Thorogood M, Hillsdon M. Remote and web 2. 0 interventions for promoting physical activity. Cochrane Database of Systematic Reviews Internet John Wiley & Sons, Ltd 2013:1-11. [doi: 10.1002/14651858.cd010395]
6. Whittaker R, McRobbie H, Bullen C, Rodgers A, Gu Y. Mobile phone-based interventions for smoking cessation. Cochrane Database Syst Rev 2016 Apr 10;4:CD006611 [FREE Full text] [doi: 10.1002/14651858.CD006611.pub4] [Medline: 27060875]
7. Cui M, Wu X, Mao J, Wang X, Nie M. T2DM self-management via smartphone applications: a systematic review and meta-analysis. PLoS One 2016;11(11):e0166718 [FREE Full text] [doi: 10.1371/journal.pone.0166718] [Medline: 27861583]
8. Chow CK, Ariyarathna N, Islam SMS, Thiagalingam A, Redfern J. mHealth in cardiovascular health care. Heart Lung Circ 2016 Aug;25(8):802-807. [doi: 10.1016/j.hlc.2016.04.009] [Medline: 27262389]
9. US smartphone ownership 2011-2019. Statista. URL: https://www.statista.com/statistics/219865/percentage-of-us-adults-who-own-a-smartphone/ [accessed 2020-09-01]
10. West JH, Belvedere LM, Andreasen R, Frandsen C, Hall PC, Crookston BT. Controlling your "app"etite: how diet and nutrition-related mobile apps lead to behavior change. JMIR Mhealth Uhealth 2017 Jul 10;5(7):e95 [FREE Full text] [doi: 10.2196/mhealth.7410] [Medline: 28694241]
11. Payne HE, Lister C, West JH, Bernhardt JM. Behavioral functionality of mobile apps in health interventions: a systematic review of the literature. JMIR Mhealth Uhealth 2015 Feb 26;3(1):e20 [FREE Full text] [doi: 10.2196/mhealth.3335] [Medline: 25803705]
12. Bailey SC, Belter LT, Pandit AU, Carpenter DM, Carlos E, Wolf MS. The availability, functionality, and quality of mobile applications supporting medication self-management. J Am Med Inform Assoc 2014;21(3):542-546 [FREE Full text] [doi: 10.1136/amiajnl-2013-002232] [Medline: 24163156]

13. Donevant SB, Estrada RD, Culley JM, Habing B, Adams SA. Exploring app features with outcomes in mHealth studies involving chronic respiratory diseases, diabetes, and hypertension: a targeted exploration of the literature. J Am Med Inform Assoc 2018 Oct 01;25(10):1407-1418 [FREE Full text] [doi: 10.1093/jamia/ocy104] [Medline: 30137383]
14. Widmer RJ, Collins NM, Collins CS, West CP, Lerman LO, Lerman A. Digital health interventions for the prevention of cardiovascular disease: a systematic review and meta-analysis. Mayo Clin Proc 2015 Apr;90(4):469-480 [FREE Full text] [doi: 10.1016/j.mayocp.2014.12.026] [Medline: 25841251]
15. Abroms LC, Boal AL, Simmens SJ, Mendel JA, Windsor RA. A randomized trial of Text2Quit: a text messaging program for smoking cessation. Am J Prev Med 2014 Sep;47(3):242-250 [FREE Full text] [doi: 10.1016/j.amepre.2014.04.010] [Medline: 24913220]
16. ASA24 dietary assessment tool. National Institutes of Health. URL: https://epi.grants.cancer.gov/asa24/ [accessed 2020-09-01]
17. Norman GJ, Kolodziejczyk JK, Adams MA, Patrick K, Marshall SJ. Fruit and vegetable intake and eating behaviors mediate the effect of a randomized text-message based weight loss program. Prev Med 2013 Jan;56(1):3-7 [FREE Full text] [doi: 10.1016/j.ypmed.2012.10.012] [Medline: 23085329]
18. Leong DP, Joseph PG, McKee M, Anand SS, Teo KK, Schwalm J, et al. Reducing the global burden of cardiovascular disease, part 2: prevention and treatment of cardiovascular disease. Circ Res 2017 Sep 01;121(6):695-710. [doi: 10.1161/CIRCRESAHA.117.311849] [Medline: 28860319]
19. Islam MN, Karim MM, Inan TT, Islam AKMN. Investigating usability of mobile health applications in Bangladesh. BMC Med Inform Decis Mak 2020 Feb 03;20(1):19 [FREE Full text] [doi: 10.1186/s12911-020-1033-3] [Medline: 32013965]
20. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. Int J Med Inform 2019 Jun;126:95-104. [doi: 10.1016/j.ijmedinf.2019.03.018] [Medline: 31029270]
21. Sheehan B, Lee Y, Rodriguez M, Tiase V, Schnall R. A comparison of usability factors of four mobile devices for accessing healthcare information by adolescents. Appl Clin Inform 2012;3(4):356-366 [FREE Full text] [doi: 10.4338/ACI-2012-06-RA-0021] [Medline: 23227134]
22. Manger bouger. Republic of France. URL: https://www.mangerbouger.fr/ [accessed 2020-09-01]
23. Haut Conseil de la Santé Publique. URL: https://www.hcsp.fr/explore.cgi/Accueil [accessed 2020-09-01]
24. Bienvenue dans une vie sans tabac. Tabac-Info-Service of France. URL: https://www.tabac-info-service.fr [accessed 2020-09-01]
25. Traitement de l'alcoolisme: Décider d'arrêter l'alcool. Alcool Info Service. URL: https://www.alcool-info-service.fr/alcool/traitement-alcoolisme/decider-arreter-alcool [accessed 2020-09-01]
26. Shneiderman B, Plaisant C. Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition). Boston: Pearson Addison Wesley; Apr 10, 2004:-321.
27. Belden J, Grayson R, Barnes J. Defining and testing EMR usability. Research Gate.: Healthcare Information and Management Systems Society (HIMSS); 2009 Jun 01. URL: https://www.researchgate.net/publication/277829258_Defining_and_Testing_EMR_Usability_Principles_and_Proposed_Methods_of_EMR_Usability_Evaluation_and_Rating [accessed 2022-01-04]
28. Gluon Documentation. URL: https://docs.gluonhq.com/ [accessed 2020-09-01]
29. Tsopra R, Jais J, Venot A, Duclos C. Comparison of two kinds of interface, based on guided navigation or usability principles, for improving the adoption of computerized decision support systems: application to the prescription of antibiotics. J Am Med Inform Assoc 2014 Feb;21(e1):e107-e116 [FREE Full text] [doi: 10.1136/amiajnl-2013-002042] [Medline: 24008427]
30. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth 2015 Mar 11;3(1):e27 [FREE Full text] [doi: 10.2196/mhealth.3422] [Medline: 25760773]
31. Stoyanov SR, Hides L, Kavanagh DJ, Wilson H. Development and validation of the user version of the Mobile Application Rating Scale (uMARS). JMIR Mhealth Uhealth 2016 Jun 10;4(2):e72 [FREE Full text] [doi: 10.2196/mhealth.5849] [Medline: 27287964]
32. Moorman M, van den Putte B. The influence of message framing, intention to quit smoking, and nicotine dependence on the persuasiveness of smoking cessation messages. Addict Behav 2008 Oct;33(10):1267-1275. [doi: 10.1016/j.addbeh.2008.05.010] [Medline: 18584971]
33. Yang DJ. The communication effects of audience situation and message framing on smoking cessation. SSRN Journal 2013 Dec 02:252-264. [doi: 10.2139/ssrn.2368006]
34. Liberati EG, Ruggiero F, Galuppo L, Gorli M, González-Lorenzo M, Maraldi M, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017 Sep 15;12(1):113 [FREE Full text] [doi: 10.1186/s13012-017-0644-2] [Medline: 28915822]
35. Miller K, Mosby D, Capan M, Kowalski R, Ratwani R, Noaiseh Y, et al. Interface, information, interaction: a narrative review of design and functional requirements for clinical decision support. J Am Med Inform Assoc 2018 May 01;25(5):585-592 [FREE Full text] [doi: 10.1093/jamia/ocx118] [Medline: 29126196]
36. Arnett DK, Blumenthal RS, Albert MA, Buroker AB, Goldberger ZD, Hahn EJ, et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: Executive Summary: A Report of the American College of

Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2019 Sep 10;74(10):1376-1414 [FREE Full text] [doi: 10.1016/j.jacc.2019.03.009] [Medline: 30894319]
37. Michie S, Wood CE, Johnston M, Abraham C, Francis JJ, Hardeman W. Behaviour change techniques: the development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). Health Technol Assess 2015 Nov;19(99):1-188 [FREE Full text] [doi: 10.3310/hta19990] [Medline: 26616119]
38. Wakefield MA, Loken B, Hornik RC. Use of mass media campaigns to change health behaviour. Lancet 2010 Oct 09;376(9748):1261-1271 [FREE Full text] [doi: 10.1016/S0140-6736(10)60809-4] [Medline: 20933263]
39. Wakefield MA, Durkin S, Spittal MJ, Siahpush M, Scollo M, Simpson JA, et al. Impact of tobacco control policies and mass media campaigns on monthly adult smoking prevalence. Am J Public Health 2008 Aug;98(8):1443-1450. [doi: 10.2105/AJPH.2007.128991] [Medline: 18556601]
40. Marcolino MS, Oliveira JAQ, D'Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The impact of mHealth interventions: systematic review of systematic reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
41. Stephens TN, Joerin A, Rauws M, Werk LN. Feasibility of pediatric obesity and prediabetes treatment support through Tess, the AI behavioral coaching chatbot. Transl Behav Med 2019 May 16;9(3):440-447. [doi: 10.1093/tbm/ibz043] [Medline: 31094445]
42. Piette JD, List J, Rana GK, Townsend W, Striplin D, Heisler M. Mobile health devices as tools for worldwide cardiovascular risk reduction and disease management. Circulation 2015 Nov 24;132(21):2012-2027 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.114.008723] [Medline: 26596977]
43. Balmain BN, Tuttle N, Bailey J, Cheng K, Duryea M, Houlihan J, et al. Using smart socks to detect step-count at slow walking speeds in healthy adults. Int J Sports Med 2019 Feb;40(2):133-138. [doi: 10.1055/a-0732-5621] [Medline: 30544268]
44. Fagherazzi G, Ravaud P. Digital diabetes: Perspectives for diabetes prevention, management and research. Diabetes Metab 2019 Sep;45(4):322-329. [doi: 10.1016/j.diabet.2018.08.012] [Medline: 30243616]
45. Kononova A, Li L, Kamp K, Bowen M, Rikard RV, Cotten S, et al. The use of wearable activity trackers among older adults: focus group study of tracker perceptions, motivators, and barriers in the maintenance stage of behavior change. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e9832 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 30950807]
46. Mercer K, Li M, Giangregorio L, Burns C, Grindrod K. Behavior change techniques present in wearable activity trackers: a critical analysis. JMIR Mhealth Uhealth 2016 Apr 27;4(2):e40 [FREE Full text] [doi: 10.2196/mhealth.4461] [Medline: 27122452]
47. Tudor Car L, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
48. Laranjo L, Dunn AG, Tong HL, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
49. Hauser-Ulrich S, Künzli H, Meier-Peterhans D, Kowatsch T. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): pilot randomized controlled trial. JMIR Mhealth Uhealth 2020 Apr 03;8(4):e15806 [FREE Full text] [doi: 10.2196/15806] [Medline: 32242820]
50. Kramer LL, Ter Stal S, Mulder BC, de Vet E, van Velsen L. Developing embodied conversational agents for coaching people in a healthy lifestyle: scoping review. J Med Internet Res 2020 Feb 06;22(2):e14058 [FREE Full text] [doi: 10.2196/14058] [Medline: 32022693]

Abbreviations

CVD: cardiovascular disease
mHealth: mobile health
uMARS: user version of the Mobile App Rating Scale

Edited by R Kukafka; submitted 11.11.20; peer-reviewed by S Donevant, E Sezgin, J Whiteley; comments to author 01.02.21; revised version received 04.03.21; accepted 25.04.21; published 20.01.22.

Please cite as:
Agher D, Sedki K, Despres S, Albinet JP, Jaulent MC, Tsopra R
Encouraging Behavior Changes and Preventing Cardiovascular Diseases Using the Prevent Connect Mobile Health App: Conception and Evaluation of App Quality
J Med Internet Res 2022;24(1):e25384
URL: https://www.jmir.org/2022/1/e25384
doi:10.2196/25384
PMID:35049508

Original Paper

Social Media Platforms Listening Study on Atopic Dermatitis: Quantitative and Qualitative Findings

Paméla Voillot[1], MSc; Brigitte Riche[2], MS; Michel Portafax[2], MSc, MS; Pierre Foulquié[1], MSc; Anaïs Gedik[1], MSc; Sébastien Barbarot[3], MD; Laurent Misery[4], PhD; Stéphane Héas[5], MS; Adel Mebarki[1], MS; Nathalie Texier[1], PharmD; Stéphane Schück[1], MD

[1]Kap Code, Paris, France
[2]Sanofi Genzyme, Gentilly, France
[3]Centre Hospitalier Universitaire de Nantes, Nantes, France
[4]Centre Hospitalier Universitaire de Brest, Brest, France
[5]Université de Rennes 2, Rennes, France

Corresponding Author:
Paméla Voillot, MSc
Kap Code
28 rue d'Enghien
Paris, 75010
France
Phone: 33 972605764
Email: pamela.voillot@kapcode.fr

Abstract

Background: Atopic dermatitis (AD) is a chronic, pruritic, inflammatory disease that occurs most frequently in children but also affects many adults. Social media have become key tools for finding and disseminating medical information.

Objective: The aims of this study were to identify the main themes of discussion, the difficulties encountered by patients with respect to AD, the impact of the pathology on quality of life (QoL; physical, psychological, social, or financial), and to study the perception of patients regarding their treatment.

Methods: A retrospective study was carried out by collecting social media posts in French language written by internet users mentioning their experience with AD, their QoL, and their treatments. Messages related to AD discomfort posted between July 1, 2010, and October 23, 2020, were extracted from French-speaking publicly available online forums. Automatic and manual extractions were implemented to create a general corpus and 2 subcorpuses depending on the level of control of the disease.

Results: A total of 33,115 messages associated with AD were included in the analysis corpus after extraction and cleaning. These messages were posted by 15,857 separate web users, most of them being women younger than 40 years. Tips to manage AD and everyday hygiene/treatments were among the most discussed topics for controlled AD subcorpus, while baby-related topics and therapeutic failure were among the most discussed topics for insufficiently controlled AD subcorpus. QoL was discussed in both subcorpuses with a higher proportion in the controlled AD subcorpus. Treatments and their perception were also discussed by web users.

Conclusions: More than just emotional or peer support, patients with AD turn to online forums to discuss their health. Our findings show the need for an intersection between social media and health care and the importance of developing new approaches such as the Atopic Dermatitis Control Tool, which is a patient-related disease severity assessment tool focused on patients with AD.

(J Med Internet Res 2022;24(1):e31140) doi:10.2196/31140

KEYWORDS

atopic dermatitis; Atopic Dermatitis Control Tool; health-related quality of life; social media use; real world; dermatology; skin disease; social media; online health information; online health; health care

Introduction

Atopic dermatitis (AD) is a chronic, pruritic, inflammatory skin disease that occurs most frequently in children but that can also affect adults. The course of the disease is relapsing, and it is frequently associated with elevated levels of serum immunoglobulin E, individual or family history of food allergies, allergic rhinitis, and asthma [1-3]. According to the World Health Organization Global Burden of Diseases initiative's data, AD ranks 15th among all nonfatal disabilities worldwide and has the highest disease burden among skin diseases as measured by disability-adjusted life-years [4]. Childhood-onset AD begins early in life, with 50% diagnosed in the first year of life and 85% by 5 years of age [1,5]. However, AD can present at any age, with adult onset reported by 26% of patients with AD [6]. Although AD often resolves during childhood, it persists through adulthood in 20%-50% of patients [7,8].

AD is associated with substantial morbidity and quality of life (QoL) impairment. There are several comorbid health problems that occur in patients with AD, aside from the cutaneous signs and symptoms. Chronic pruritus and inflammation can lead to sleep disturbances and mental health symptoms, which are not mutually exclusive. AD may also predispose to a higher risk of other atopic disorders, including asthma and allergic rhinitis [9]. Persons with AD appear to be at higher risk for multiple neuropsychiatric disorders, including depression, anxiety, attention-deficit hyperactivity disorder, speech disorders in childhood, headaches, and seizures [10]. There is also a multifactorial association of AD with osteoporosis, bone and joint injuries, infections, and fractures [11-14].

Clinical presentation and severity of AD vary widely, and diagnosis is not always straightforward, especially in adults [15]. Treatment of AD follows a multifaceted, stepwise approach that is tailored according to disease severity [1]. For all patients, basic management and flare prevention consist of good skin care practices (daily showers or baths) followed immediately by the application of emollients and moisturizers, with avoidance of triggers such as irritants; aero or food allergens; and extremes of heat, cold, or humidity [9]. In mild AD, treatment involves, as needed, use of low- to mid-potency topical corticosteroids or topical calcineurin inhibitors during flares. Patients with frequent flares may benefit from proactive application of topical anti-inflammatory therapies twice a week to the most troublesome areas. Patients with severe disease often present significant treatment challenges. Systemic therapies are usually required for severe AD but have varying degrees of success and can be associated with side effect profiles that require counseling and close monitoring. Phototherapy has been shown to have success in treating moderate-to-severe AD, but several factors can limit its utility and efficacy including cost and access. New therapies are targeting specific pathways relevant for AD and many others are in development. An array of topical, oral, and injectable therapies targeting specific disease pathways in AD are in development for pediatric and adult populations [16]. Dupilumab (Dupixent) is the current first-line systemic agent for adults and children with moderate-to-severe, treatment-resistant AD. It is the only biologic treatment approved from 6 years of age. There are several emerging therapies currently in Phase III clinical trials such as JAK1 and JAK2 inhibitors or other biologic treatments (monoclonal antibodies or NK-1R antagonists) [16,17].

Social media is one of the most rapid and impactful ways of obtaining and delivering information in the modern era. In general, social media refers to forms of electronic communication (such as websites for social networking and microblogging) through which users create online communities to share information, opinions, personal messages, photos, videos, and other contents within internet applications [18]. Social media provides a readily accessible means to promote user-generated content, broaden interpersonal connections, and encourage social collaboration.

Social media has gained unprecedented worldwide popularity over the last 2 decades. It is estimated that there are currently over 2.3 billion active social media users internationally and this number is growing by approximately 1 million new users every day [19]. Unsurprisingly, therefore, the use of social media to find, exchange, and discuss health information is growing at an unprecedented rate. Anyone with access to the internet can post or read information on a (social media) site. This means that it is directly accessible to patients, their family and friends, and all health care providers. They provide access to active communities of health care professionals and fellow patients, with whom they can share information and their experiences, raise awareness of their concerns, learn about their conditions and health care opportunities, and find support [18].

Artificial intelligence (AI) tools are strongly related to data mining and AI is nowadays ranked among the top-10 technology [20]. Despite their limitations, AI tools and techniques that are still in their infancy already offer substantial benefits in providing in-depth knowledge on individuals' health and predicting population health risks, and their use for medicine and public health is likely to increase substantially in the near future [21].

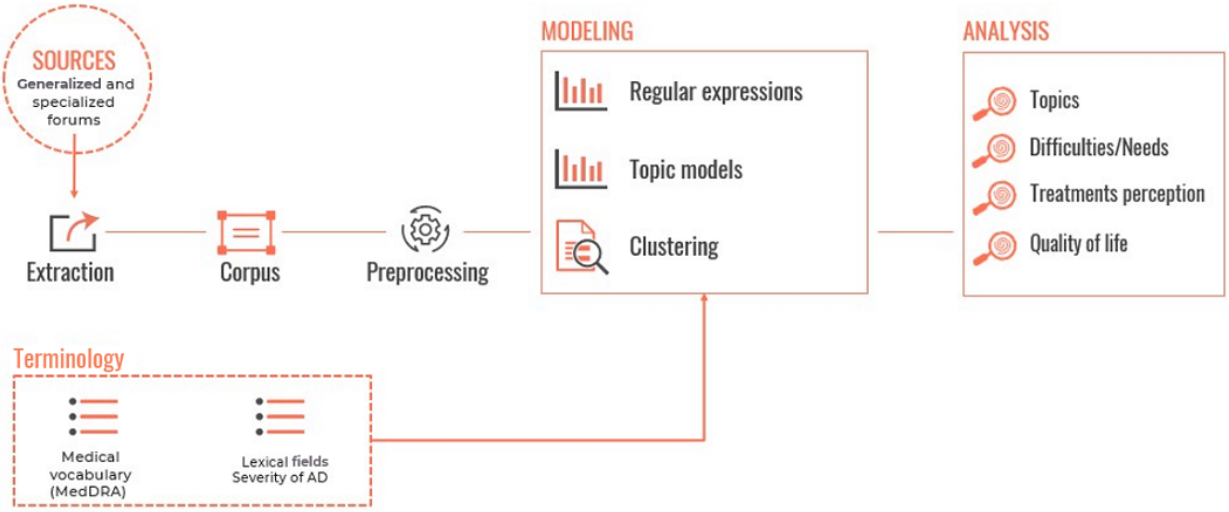
The purpose of this retrospective study was to better understand how patients suffering from AD perceive their QoL and their treatments, based on the assessment of web-based social media posts, which we considered as a real-life source of health information. In this study, we aimed to (1) identify the main themes for discussion; (2) identify the difficulties encountered by patients with respect to AD and the impact of the disease on QoL (physical, psychological, social, or financial); and (3) study the perception of patients regarding their treatment.

Methods

Study Design and Population

This was a noninterventional retrospective study using a text mining approach to retrieve information from social media posts (data available in the public domain) written by French language internet users between July 1, 2010, and October 23, 2020 (Figure 1). The study was conducted in 2 phases: data collection using the published Detec't web crawler [22,23] developed by Kap Code to collect AD-related posts, and a quantitative/qualitative analysis to identify trends and characterize key themes discussed by French-speaking users.

Figure 1. Study framework. AD: atopic dermatitis.

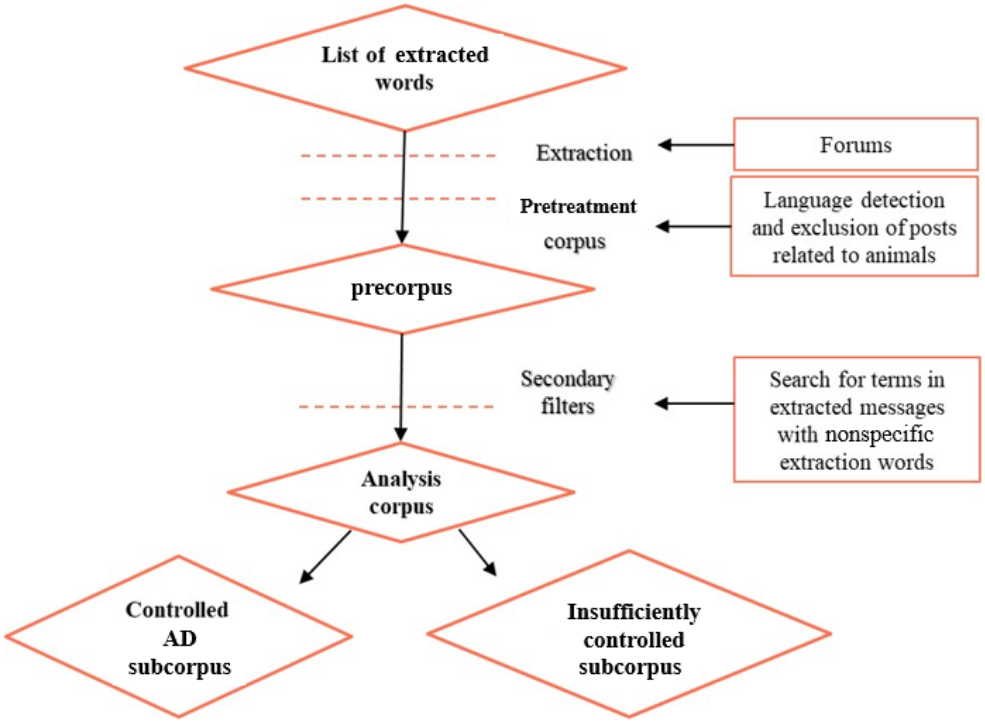


Data Extraction

A web crawler is an engine that browses through hyperlinks and stores them for future download of associated web pages (identified by the visited hyperlinks) [24]. Scraping of messages was performed according to the HTML structure of each forum. All discussions containing at least one of the keywords or one of their synonyms were automatically retrieved with all the associated metadata, deidentified, and cleaned (signature and quote withdrawal) before being stored in a study-specific database. A list of the keywords used for message retrieval is detailed in Multimedia Appendix 1 and a list of secondary filters on nonspecific extraction words is detailed in Multimedia Appendix 2. These terms were searched in extracted messages with extraction words not specific to AD. If one of these words was found in the post, then the message was kept in the corpus.

The analysis corpus consisted of the corpus cleaned after the removal of messages containing predetermined keywords written in a language other than French, posts containing animal-related vocabulary, and messages containing at least one of the study-specific inclusion words listed in Multimedia Appendix 2. The analysis corpus was then divided into 2 subcorpuses: 1 subcorpus for *controlled AD* and 1 subcorpus for *insufficiently controlled AD* (Figure 2). The lexical fields of both subcorpuses were realized based on the Atopic Dermatitis Control Tool (ADCT) questionnaire and completed with real expressions of internet users [22].

Figure 2. Flowchart presenting the steps for creating analysis corpus and the 2 subcorpuses. AD: atopic dermatitis.



Data Analysis

Age and Gender

Web users' gender was determined through the identification of regular expressions for each gender. First names and gender-associated suffix and prefix were first searched in the username. Then, the content of all available messages was screened for gender-specific lexical fields and gender agreements of adjectives and verbs. In the end, a score was computed for each gender and a prediction for the user was obtained by comparing them.

Ages of web users were identified by a twofold method. First, regular expressions of age were identified in the entire user post. If no expressions were found, then a machine learning model predicted the age based on several features. Among them, the model considered syntactic aspects of posts as well as expressed feeling and the source on which the users express themselves.

Topic Model

A topic model was applied to identify the themes addressed in the messages. Topic models consist of a text mining approach aiming to automatically identify the abstract themes addressed in a collection of documents. Such models are based on the hypothesis that each document in the corpus corresponds to a distribution of several topics. A Biterm Topic Model (BTM) was used to identify the topics without prior knowledge. A topic is defined as a subject of discussion, which amounts to tokens that frequently appear together in a corpus. The BTM considers the whole corpus as a mixture of topics, where each co-occurring pair of tokens (the biterm) is drawn from a specific topic independently and modeled topics are probability distributions over the biterms.

As topics are probability distributions over tokens of the study corpus, they can be characterized by the highest per-topic probability tokens. Weighting these probabilities through term-frequency inverse document frequency (TF-IDF) weighting allows to allocate a higher importance to topic-specific tokens. In this case, the per-topic probability of a token was weighted by the inverse of the probabilities of this token in other topics. For each topic, tokens were ranked from the highest to the lowest weighted probabilities TF-IDF value in this topic. The first 15 tokens were designated as the set of characteristic tokens and used to manually name the topic. A topic model was applied to each of the subcorpuses.

Health-Related Quality of Life

A health-related quality of life (HRQoL) algorithm was applied to identify and qualify expressed impact of the disease or treatments on the QoL of a patient. Types of impact were defined according to HRQoL survey categories. The algorithm was twofold: it indicates if an impact is expressed and, thanks to 5 specific models, it indicates the nature of the impact: physical, psychic, activity-related, relational, or financial. Features involved in the model describe expressed emotions, grammar, conjugation, and lexical fields of HRQoL-related features. The HRQoL algorithm was applied to each of the subcorpuses.

Drug Intake

To identify drug intake by the author of the message, the first step consisted of identifying the treatments cited in the 2 subcorpuses of messages. These were detected from the Detec't database, which contains around 2500 molecules and drugs. After that, a prediction was made the simultaneous presence of the drug name and its associated adverse reaction in the same message (ie, a product–message couple). The model bases its prediction on regular forms expressing the intake of a treatment and their distance to the drug mention. Expressions were scored according to their probability of implying a treatment intake.

Feeling Analysis

In both subcorpuses, the content of the drug intake messages was analyzed (sentiment analysis) using Microsoft Azure Cognitive Services. Sentiment analysis is part of text analytics and can detect the level of positive or negative sentiment of input text using a confidence score across a variety of languages. This machine learning algorithm assigns a sentiment score to each sentence of a message, and then calculates the overall score to assign the message to a category (positive, negative, mixed, or neutral).

The sentiment analysis algorithm was applied to each of the drug intake messages' subcorpuses. For drugs with few identified intakes, sentiment analysis was applied to treatment citation instead of treatment intake: this was the case for Dupixent for example.

The qualitative content of the positive and negative feelings in the messages was characterized by a manual identification with a review of posts detected with a treatment taken by the web users. The percentage of messages was calculated according to the total number of messages identified with terms related to treatments taken.

Results

Description of the Population and Posts

Overview

After cleaning and formatting, the analysis corpus contained a total of 33,115 messages corresponding to 15,857 different web users, with a median of 2.1 posts per patient (SD 9.55) (Figure 3). The *controlled AD* subcorpus contained 9454 posts corresponding to 4875 web users and the *insufficiently controlledAD* subcorpus contained about twice as many messages and web users with 17,384 posts from 9292 web users who published fewer messages than the other subcorpus (median of 1.87 posts per patient versus 1.94).

For both subgroups, extracted data mostly came from Baby Center [25] with 889/9454 (9.40%) posts for the *controlled AD* subcorpus and 2475/17,384 (14.24%) posts for the *insufficiently controlled AD* subcorpus (Table 1). The most frequently identified keyword was "Eczema" (7717/9454 posts for the *controlled AD* subcorpus and 13,039/17,384 posts for the *insufficiently controlled AD* subcorpus) followed by "Dermatite" (1146/9454 posts for the *controlled AD* subcorpus and 2386/17,384 posts for the *insufficiently controlled AD* subcorpus).

Figure 3. Methodology. AD: atopic dermatitis.

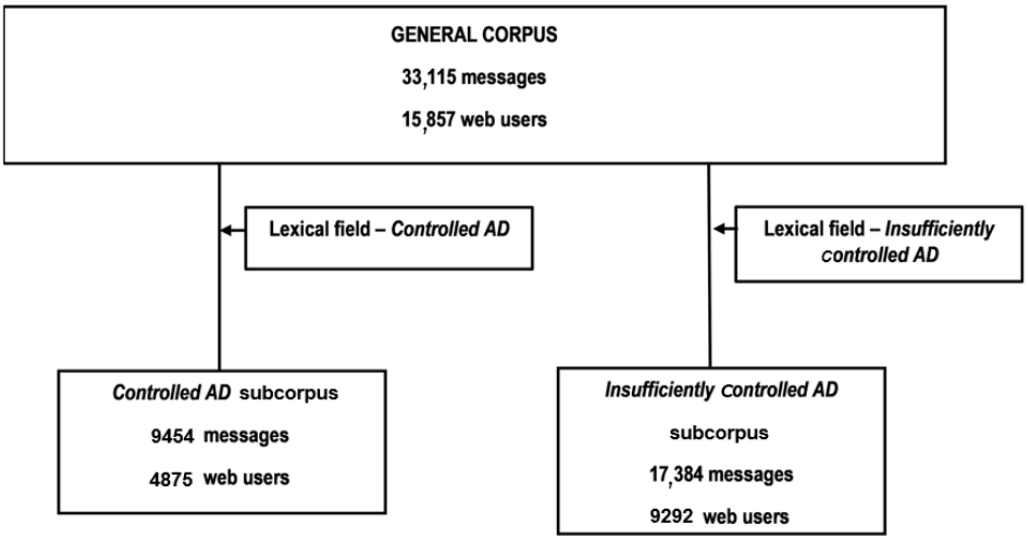


Table 1. Top 10 sites.

Controlled AD[a]		Insufficiently controlled AD	
Forums	Posts, n	Forums	Posts, n
Baby Center	889	Baby Center	2475
Aufeminin	307	Twitter	2431
L'Appart' des Spasmos	230	L'Appart' des Spasmos	1288
Amazon	220	Aufeminin	893
Mamandco	206	Mamandco	385
Twitter	128	Boursorama	282
e-Sante	109	Amazon	228
Forum Melodie	88	Mjeuxvideo	197
Club Beaute Addict	61	Forum manucure	126
Journal des Femmes	38	BourseReflex	121

[a]AD: atopic dermatitis.

We observed fluctuations in the evolution of the volume of messages between July 2010 and October 2020 (Figure 4) due to either seasonality or publications of scientific results related to AD. An apparent recurrent increase in the number of messages was observed from November to July that could be related to the increase in dust mites' allergies in winter and the discomfort associated with sweating in summer [26,27]. From February to December 2015 there was a significant increase in the number of posts, which seems to be related to the publication of different scientific articles especially concerning the results of tofacitinib citrate, an oral Janus kinase inhibitor. Another increase in messages was also seen in 2018 that can be associated with the availability of Dupixent (dupilumab) in hospitals.

In both subcorpuses, most users were women (3486/4869, 71.60%) with an average age of about 38 years and 19.41% (945/4869) appeared with undetermined gender (Figure 5). This was consistent with the results of many studies that point out that women express more personal issues in social networks [28,29]. In the present case, the topic of the disease was much discussed by mothers of children with AD who tended to focus on the appearance of their child's skin.

As shown in Figure 6, patients with AD are increasingly active online and use social media to participate, share their concerns, and express themselves freely regarding their disease. This trend was comparable to that of patients with psoriasis, even if the number of messages was lower for AD that was not sufficiently prioritized. By contrast, the comparison with the cancer disease area (1% of the messages) clearly highlighted the lack of specific forums or associations for AD compared with the cancer field.

Figure 4. Fluctuations in the evolution of the volume of messages between July 2010 and October 2020. AD: atopic dermatitis.

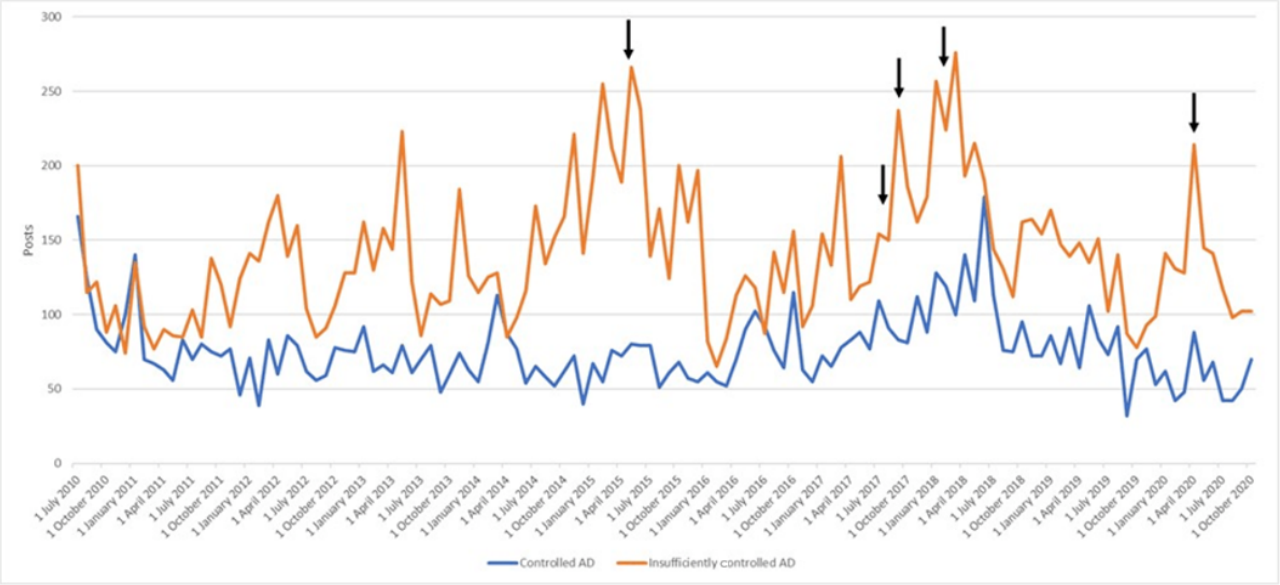


Figure 5. Age and gender distribution among (A) controlled AD subcorpus and (B) insufficiently controlled subcorpus. AD: atopic dermatitis.

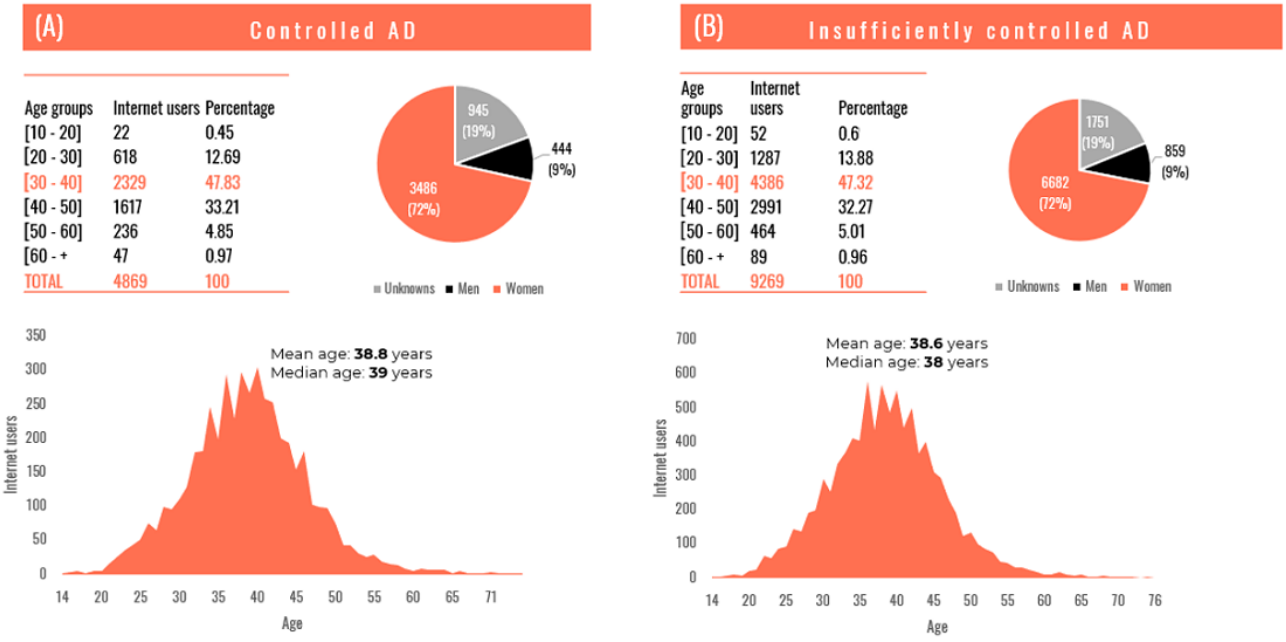
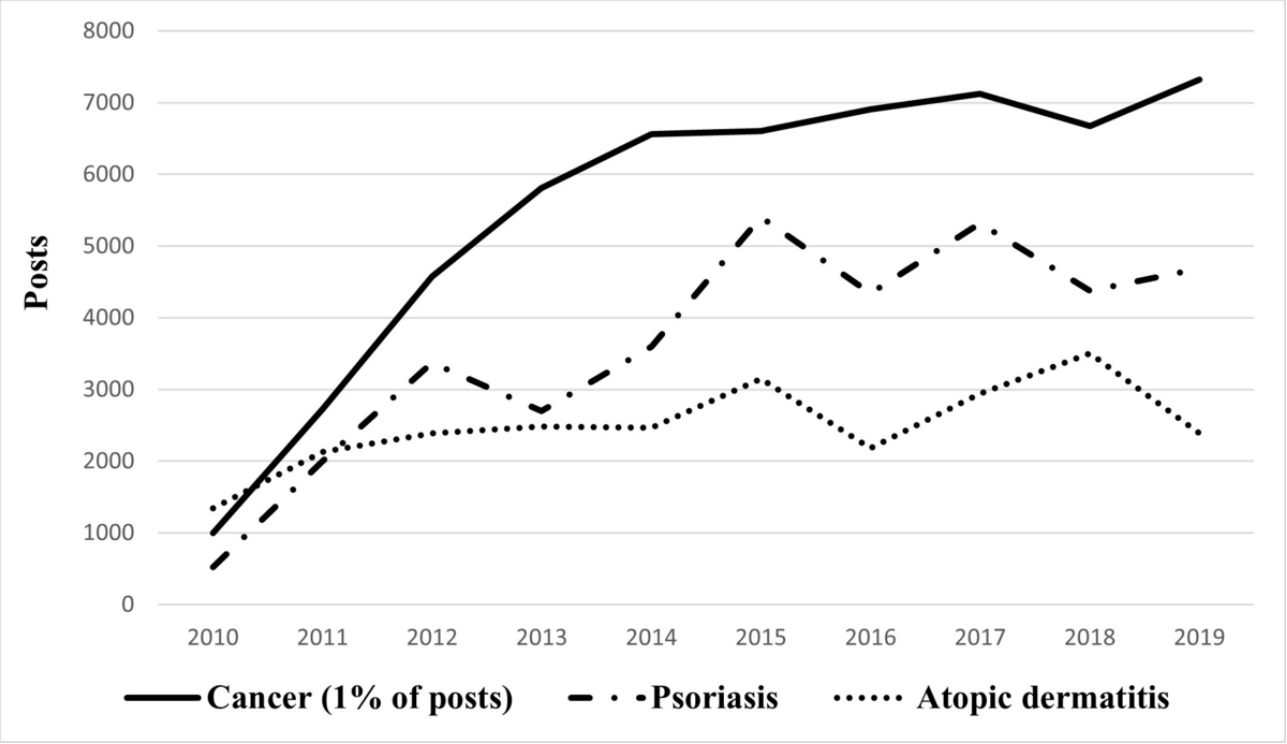


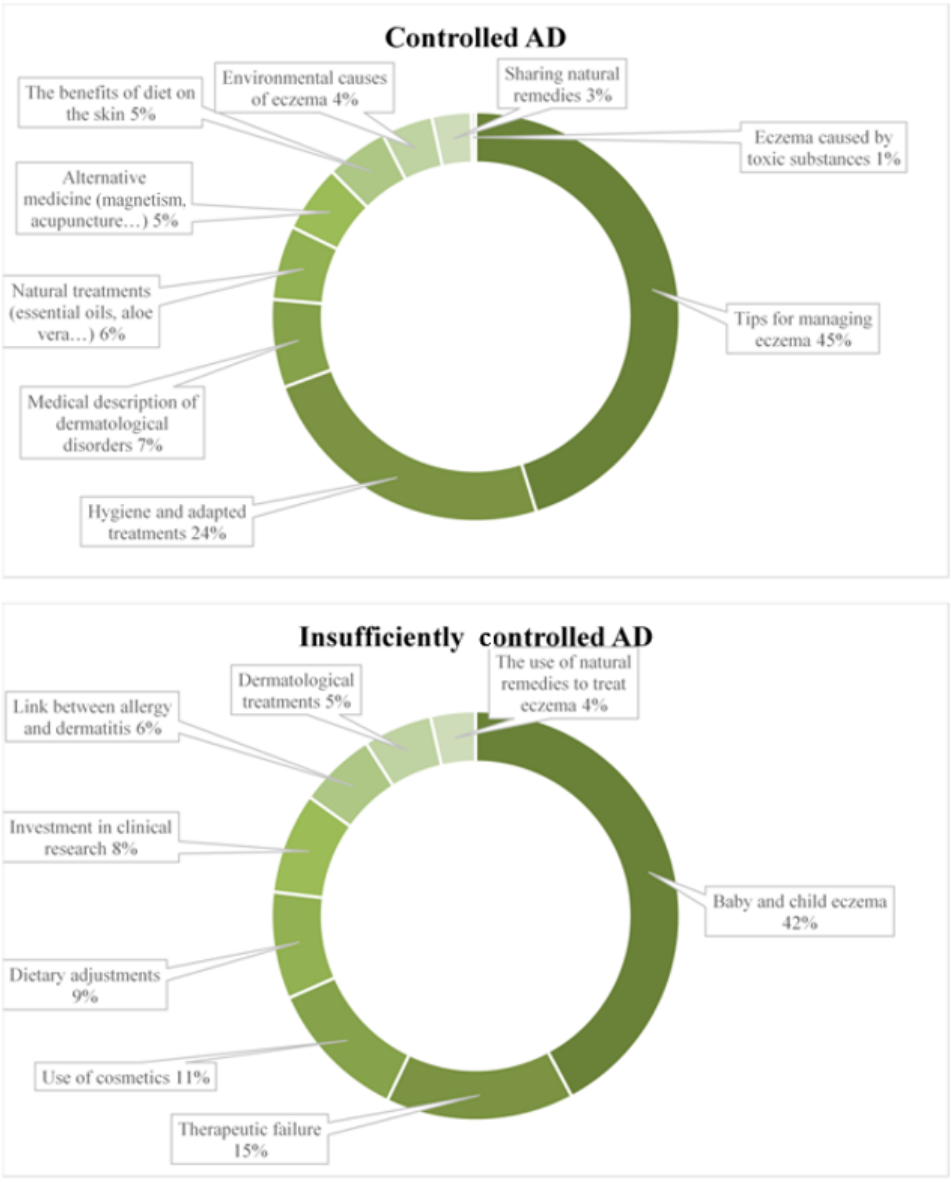
Figure 6. Trend over time of the number of posts concerning atopic dermatitis compared with other diseases such as cancer and psoriasis.



Discussion Themes and Topics

Basically, the topics around AD were roughly the same between the 2 subcorpuses and the observed differences were due to the stage of the disease. Patients in the *controlled AD* subcorpus had more perspective and experience, whereas patients in the *insufficiently controlled AD* subcorpus lacked experience and were looking for information. As shown in Figure 7, the largest category of topics was related to "the exchange of tips to relieve eczema" (2864/6339, 45.18%), followed closely by "hygiene and care adapted with the disease" (1535/6339, 24.22%) for the *controlled AD* subcorpus. For the *insufficiently controlled AD* subcorpus, 42.19% (6351/15,055) of the discussions were about "eczema in babies and children" with most messages related to discussions between parents who are concerned about their child's eczema and who ask questions to other web users. The second most discussed topic, with 14.93% (2247/15,055) of messages in the *insufficiently controlled AD* subcorpus, was the therapeutic failure experienced by web users. Their concerns reflected the poor knowledge of their disease and treatments. They expressed their frustration and lack of satisfaction about the treatments and solutions they have tested to reduce their symptoms. In the *controlled AD* subcorpus, by contrast, web users exchange information about natural remedies and alternative medicines to treat their AD. They expressed their satisfaction with the use of essential oils, food supplements, and other nonmedicinal methods and the effectiveness of these solutions. They used social networks to provide advice based on their own experience.

Figure 7. Distribution of posts by overall topic category. AD: atopic dermatitis.



Quality of Life

The QoL is an important tool for evaluating the effect of a disease as well as effects of treatment interventions. In this study, the impact of AD on the QoL of individuals was assessed by monitoring physical well-being, social well-being, material well-being, emotional well-being, and development/activity.

The impact of the disease on QoL was discussed by 70.32% (3428/4875) of patients in the *controlled AD* subcorpus (Figure 8A) and 49.05% (4558/9292) in the *insufficiently controlled AD* subcorpus (Figure 8B). This lower proportion can be explained by the heavy burden that AD places on patients with *insufficiently controlled AD*, particularly those who search for a solution to improve their health rather than their QoL. Furthermore, this trend can be related to the lack of awareness about their disease compared with patients from the *controlled AD* subcorpus, who have more knowledge about AD and therefore speak about their lived experience.

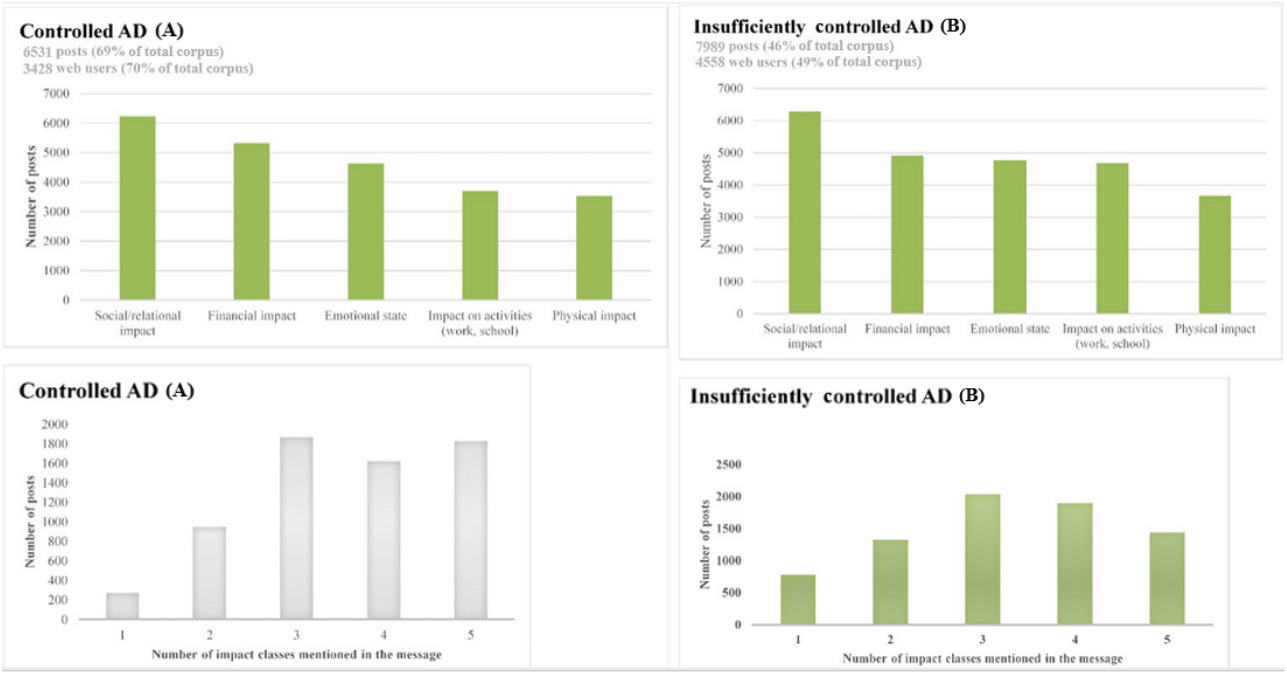
Patients in both subcorpuses seem to report the same impacts in their messages, but with a time lag that corresponds to the evolution of the knowledge of their disease.

Social functioning is affected by AD and is the first self-reported burden in both subcorpuses. Social isolation can be seen in young children with AD because adults and other children avoid interacting with them. The main effect of AD on the lives of adult patients stems from embarrassment and not wanting to be seen in public. Furthermore, some web users reported that they have been teased or bullied because of their AD. This trend was also reported by the International Study on Life with Atopic Eczema (ISOLATE) that found major impacts of AD on self-esteem [30].

The impacts that follows are the emotional impact (behavioral problems, irritability, crying, problems with treatments), the impact on activities (interference with activities such as bathing, swimming, and playing; clothing restrictions; decreased productivity at work; absenteeism), the impact on physical health (including itching, scratching, pain), and the economic

impact (over-the-counter pharmacy costs among others). Independent of the corpus, there were on average 3 different impacts per message, which show the importance of this disease in everyday life.

Figure 8. (A) Impact of the disease on controlled AD subcorpus. (B) Impact of the disease on insufficiently controlled AD subcorpus. AD: atopic dermatitis.



Treatments

AD treatment is targeted at both the disease and its symptoms. Among the 9454 messages retrieved from users in the *controlled AD* subcorpus, 1229 concerned discussions on the different treatments taken by the patients (ie, 13%). A focus was made on the top 10 and it appeared that web users in the *controlled AD* subcorpus were mainly focused on cortisone/corticosteroids and anti-inflammatory therapy. In addition, patients were prescribed antihistamines to decrease the itch–scratch cycle and antibiotics if the skin becomes superinfected (Table 2, data pertaining to "Controlled AD subcorpus"). Among the 17,384 messages retrieved from users in the *insufficiently controlled AD* subcorpus, 1677 concerned discussions on the different treatments taken by the patients (ie, 9.65%). Topically applied corticosteroids and emollients, moisturizers, and bath additives were the mainstay of therapy used by web users in the *insufficiently controlled AD* subcorpus followed by anti-inflammatory therapy to control pruritus and antihistamines as a therapeutic adjunct to alleviate it. In this subcorpus, treatments for acute phases such as topical corticosteroids, for example (Diprosone), were also discussed by patients from the *insufficiently controlled AD* subcorpus (Table 2, data pertaining to "Insufficiently controlled AD subcorpus").

Then, the perception of treatments was found in 63.30% of the messages related to treatments taken by users in the *controlled AD* subcorpus (778/1229 messages) and in 72.09% of the messages dealing with treatments taken by users in the *insufficiently controlled AD* subcorpus (1209/1677 messages). In both subcorpuses, the most expressed treatment perception was cortisone.

Mixed perception was found in 57% (47/83) of the *controlled AD* and 66% (114/173) of the *insufficiently controlled AD* with some concerns on the risks/side effects of cortisone emphasized by its efficacy. The perception was mostly negative (21/83, 25%) for *controlled AD* and 23% (39/173) for *insufficiently controlled AD* subcorpuses and positive in only 8% (7/83) of web users in the *controlled AD* and in 7% (12/173) of web users in the *insufficiently controlled AD* (Figure 9) subcorpuses. It is to note that in the lower panel, Dupixent is not a medication taken by patients with *insufficiently controlled AD* but only cited in their messages on which an analysis of perception was applied.

Concerning cortisone, the *controlled AD* subcorpus and the *insufficiently controlled AD* subcorpus expressed safety concerns associated with their long-term use with negative feelings and beliefs (25% [21/83] and 23% [39/173] of the messages, respectively), whereas 8% (7/83) of users in the *controlled AD* subcorpus and 7% (12/173) of those in the *insufficiently controlled AD* subcorpus felt a positive perception following the effectiveness they had experienced. Positive and negative perceptions were quite counter-balanced between side effects and satisfaction for the texture and the use as a preventive treatment in patients with *controlled AD* using Dexeryl (9% [2/22] of positive and negative messages), whereas patients from the *insufficiently controlled AD* subcorpus reported more concerns about safety (10/31, 32%). The ineffectiveness of anti-inflammatories to calm the flare-ups was a concern for 19% (4/21) of users in the *controlled AD* subcorpus and 33% (6/18) of users in the *insufficiently controlled AD* subcorpus, whereas only 5% (1/21) and 17% (3/18), respectively, had a positive efficacy feeling. For Dupixent, there was an absence of messages from internet users who have taken or are aware of the treatment,

and this may be due to its restricted prescription by dermatologists or directly from hospital and because it is not prescribed as first intention AD therapy. The only messages found were from patients seeking information mainly on the conditions of access (13/165, 7.9%) and on the effectiveness of the drug (13/165, 7.9%; Figure 10).

Table 2. Top 10 treatments used.

Controlled AD[a] subcorpus		Insufficiently controlled AD subcorpus	
Treatments taken (top 10)	Web users, n	Treatments taken (top 10)	Web users, n
Cortisone/corticoids	131	Cortisone/corticoids	222
Anti-inflammatories	32	Liniment	40
Dexeryl	28	Dexeryl	39
Calendula	28	Anti-inflammatories	26
Liniment	26	Calendula	26
Antibiotics	26	Antibiotics	17
Vitamins	12	Diprosone	16
Atarax	8	Bepanthen	11
Bepanthen	6	Aerius	10
Aerius	5	Vitamins	10

[a]AD: atopic dermatitis.

Figure 9. Treatments' perception. (A) Controlled AD subcorpus. (B) Insufficiently Controlled AD subcorpus. AD: atopic dermatitis.

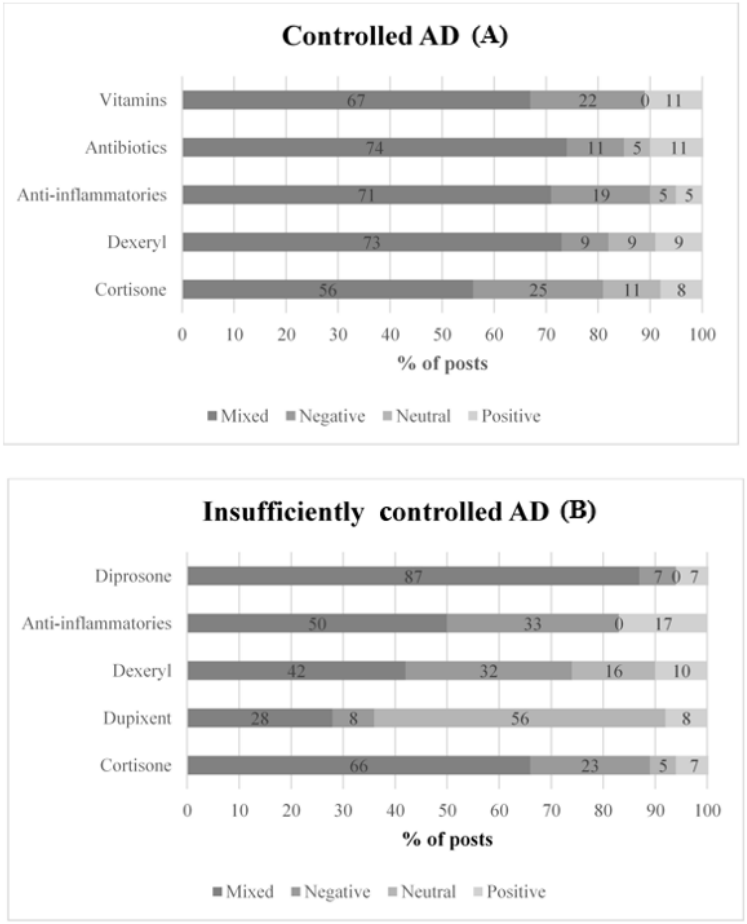


Figure 10. Treatments' perception. (A) Controlled AD subcorpus. (B) Insufficiently controlled AD subcorpus. AD: atopic dermatitis; AE: adverse event.

Discussion

Principal Findings

This retrospective study aimed at assessing how patients with AD or their parents perceive their QoL and their treatments following the analysis of web forums in France over the last 10 years. AD is increasingly discussed in French web forums as shown in this study (Figure 6), with most users being women between the ages of 30 and 40 (Figure 5).

We were able to segment web users into 2 subcorpuses corresponding to patients with *controlled AD* and patients with *insufficiently controlled AD*. These subcorpuses are characterized by similarities in the expressions. Web users from the *controlled AD* subcorpus and those from the *insufficiently controlled AD* subcorpus have almost the same preoccupations but we observed a time lag due in part to the stage and the knowledge of their disease. The main topics were the lack of information and solutions leading to some sort of therapeutic wandering, the high impact of AD on the QoL and the well-being of patients, and the dissatisfaction with the available treatments.

Limitations of the Study

Given the inherent observational nature of social media data, their analyses are subject to many limitations.

Selection bias was the first limitation because analyses were restricted to French data sources, and social media users and nonusers may differ. Thus, results are not generalizable at a worldwide scale. Furthermore, using social media to analyze patients' reactions excludes patients who do not have access to the internet or who are not familiar with the use of online discussions.

Extraction bias was the second limitation. Keyword selection in social media studies can induce varying levels of extraction bias.

Another limitation of using social media is that complete information about individual cases may be harder to obtain, unlike traditional epidemiological studies. There is also the problem of discovering demographic information—only limited or no information regarding individual user demographics may be available. Demographic information such as age and race need to be determined via automated techniques.

Health misinformation is significant on social media; nevertheless, recent reviews showed that misinformation or fake information related to health is most prevalent in studies dealing with the safety of tobacco, vaccines, and drugs such as opioids and marijuana. Finally, the lowest levels of misinformation were observed in studies evaluating medical treatments [31,32]. Nevertheless, results from Pulido et al [33] indicated that messages focused on fake health information are mostly aggressive and that messages with evidence of social impact overcome fake information.

Finally, social media represent an ideal place where patients can freely and spontaneously discuss their experiences with their therapy, thus providing valuable information on their QoL. However, this observation should be interpreted cautiously, because social media data may include a higher frequency of erroneous information, and patients posting on social media forums may not be representative of the wider patient population.

Implications and Future Research

Messages published on social networks should be integrated into the assessment of patient's QoL, as they can help to characterize the patient's experience in a more individualized and spontaneous way. Furthermore, it seems important to explore the specific areas of QoL of patients with AD and potentially enrich the existing standard questionnaires with new elements more relevant for these patients in their daily confrontation with disease and treatment.

To the extent that the most dominant topics can be interpreted as unmet informational needs, our study highlights the refinement of practical implications such as the improvement of available tools, and further analysis on the perception of treatments and the evolution of the QoL under treatment. We recommend to carry out a study segmented by year to see the

evolution of the different subjects (QoL, topics, treatments) over the years and to observe the global evolutions.

Conclusion

Social media listening offers the opportunity to consider behavior and interactions that are difficult to assess through traditional research methods. Millions of microblogs act as online communities, dealing with topics from the impact of AD on QoL through the lack of information on possible treatments and their effectiveness. Of particular interest is the exponential growth in recent years of patient support groups, and the high potential of users disseminating materials and opinions relating to AD through their posts. These forums are increasingly popular and have become an additional source of evidence, therapy, or support for the patients. Our study illustrates the current situation and the evolution over time of AD on social media platforms. Discussions on AD are becoming a frequent purpose on social media but they are still in contrast to the high frequency reported for the use of social media in research with patients with psoriasis.

All these data highlight the importance of clearly defining the role and limitations of these platforms for orienting future information campaigns and developing new models such as the ADCT centered on patients with AD [34]. This validated 6-item ADCT facilitates patient–physician communication on disease control. Such a tool might better inform health care professionals and patients with individualized measures covering QoL more in depth than existing standard questionnaires.

Acknowledgments

The medical writing support was provided by Nathalie Baudry of the microenterprise NB-Pharmaconsult. The study and the medical writing support have been funded by Sanofi.

Conflicts of Interest

BR and MP have disclosed that they are employees of Sanofi Genzyme, and may hold shares or stock options in the company. PV, PF, AG, AM, NT, and SS have disclosed that they are employees of Kap Code, a CRO, having a contract with Sanofi Genzyme to carry out this study. SB works with Almirall, Sanofi-Genzyme, Abbvie, Novartis, Janssen, Leo-Pharma, Pfizer, Eli Lilly, UCB Pharma, and Chiesi.

Multimedia Appendix 1
List of keywords used for the extraction of messages.
[DOCX File , 23 KB - jmir_v24i1e31140_app1.docx]

Multimedia Appendix 2
Number of extracted messages and associated number of web users per data source.
[DOCX File , 22 KB - jmir_v24i1e31140_app2.docx]

References

1. Eichenfield LF, Tom WL, Berger TG, Krol A, Paller AS, Schwarzenberger K, et al. Guidelines of care for the management of atopic dermatitis: section 2. Management and treatment of atopic dermatitis with topical therapies. J Am Acad Dermatol 2014 Jul;71(1):116-132 [FREE Full text] [doi: 10.1016/j.jaad.2014.03.023] [Medline: 24813302]
2. Berke R, Singh A, Guralnick M. Atopic dermatitis: an overview. Am Fam Physician 2012 Jul 01;86(1):35-42 [FREE Full text] [Medline: 22962911]
3. Drucker AM, Wang AR, Li W, Sevetson E, Block JK, Qureshi AA. The Burden of Atopic Dermatitis: Summary of a Report for the National Eczema Association. J Invest Dermatol 2017 Jan;137(1):26-30 [FREE Full text] [doi: 10.1016/j.jid.2016.07.012] [Medline: 27616422]
4. Laughter M, Maymone M, Mashayekhi S, Arents B, Karimkhani C, Langan S, et al. The global burden of atopic dermatitis: lessons from the Global Burden of Disease Study 1990-2017. Br J Dermatol 2021 Feb 29;184(2):304-309. [doi: 10.1111/bjd.19580] [Medline: 33006135]
5. Hanifin JM, Reed ML, Eczema Prevalence and Impact Working Group. A population-based survey of eczema prevalence in the United States. Dermatitis 2007 Jun;18(2):82-91. [doi: 10.2310/6620.2007.06034] [Medline: 17498413]
6. Lee HH, Patel KR, Singam V, Rastogi S, Silverberg JI. A systematic review and meta-analysis of the prevalence and phenotype of adult-onset atopic dermatitis. J Am Acad Dermatol 2019 Jun;80(6):1526-1532.e7. [doi: 10.1016/j.jaad.2018.05.1241] [Medline: 29864464]
7. Silverberg JI, Gelfand JM, Margolis DJ, Boguniewicz M, Fonacier L, Grayson MH, et al. Patient burden and quality of life in atopic dermatitis in US adults: A population-based cross-sectional study. Ann Allergy Asthma Immunol 2018 Sep;121(3):340-347 [FREE Full text] [doi: 10.1016/j.anai.2018.07.006] [Medline: 30025911]
8. Margolis JS, Abuabara K, Bilker W, Hoffstad O, Margolis DJ. Persistence of mild to moderate atopic dermatitis. JAMA Dermatol 2014 Jun 01;150(6):593-600 [FREE Full text] [doi: 10.1001/jamadermatol.2013.10271] [Medline: 24696036]

9. Silverberg JI. Selected comorbidities of atopic dermatitis: Atopy, neuropsychiatric, and musculoskeletal disorders. Clin Dermatol 2017 Jul;35(4):360-366 [FREE Full text] [doi: 10.1016/j.clindermatol.2017.03.008] [Medline: 28709566]
10. Kage P, Zarnowski J, Simon J, Treudler R. Atopic dermatitis and psychosocial comorbidities – What's new? ALS 2020;4(01):86-96. [doi: 10.5414/alx02174e]
11. Silverberg JI. Association between childhood atopic dermatitis, malnutrition, and low bone mineral density: a US population-based study. Pediatr Allergy Immunol 2015 Feb 13;26(1):54-61. [doi: 10.1111/pai.12315] [Medline: 25443466]
12. Garg NK, Silverberg JI. Eczema is associated with osteoporosis and fractures in adults: a US population-based study. J Allergy Clin Immunol 2015 Apr;135(4):1085-1087.e2. [doi: 10.1016/j.jaci.2014.10.043] [Medline: 25512080]
13. Narala S, Hata TR. Adult Atopic Dermatitis with Comorbid Atopic Disease is Associated with Increased Risk of Infections: A Population-Based Cross-Sectional Study. Dermatol Ther (Heidelb) 2017 Mar 30;7(1):111-121 [FREE Full text] [doi: 10.1007/s13555-017-0172-7] [Medline: 28138890]
14. Narla S, Silverberg JI. Association between atopic dermatitis and serious cutaneous, multiorgan and systemic infections in US adults. Ann Allergy Asthma Immunol 2018 Jan;120(1):66-72.e11 [FREE Full text] [doi: 10.1016/j.anai.2017.10.019] [Medline: 29273131]
15. Silvestre Salvador J, Romero-Pérez D, Encabo-Durán B. Atopic Dermatitis in Adults: A Diagnostic Challenge. J Investig Allergol Clin Immunol 2017 Apr 10;27(2):78-88 [FREE Full text] [doi: 10.18176/jiaci.0138] [Medline: 28071589]
16. Johnson BB, Franco AI, Beck LA, Prezzano JC. Treatment-resistant atopic dermatitis: challenges and solutions. Clin Cosmet Investig Dermatol 2019;12:181-192 [FREE Full text] [doi: 10.2147/CCID.S163814] [Medline: 30962700]
17. European Medicines Agency. Dupixent: Summary of product characteristics. EMA Dupixent Product Information. URL: https://www.ema.europa.eu/en/documents/product-information/dupixent-epar-product-information_en.pdf [accessed 2021-02-10]
18. von Muhlen M, Ohno-Machado L. Reviewing social media use by clinicians. J Am Med Inform Assoc 2012 Sep 01;19(5):777-781 [FREE Full text] [doi: 10.1136/amiajnl-2012-000990] [Medline: 22759618]
19. Donegan K, Ovelgonne H, Flores G, Fuglerud P, Georgescu A. Social media and mHealth data. HMA Subgroup Report. URL: https://www.ema.europa.eu/en/documents/report/social-media-m-health-data-subgroup-report_en.pdf [accessed 2021-02-10]
20. Datapine Website. URL: https://www.datapine.com/blog/technology-buzzwords/ [accessed 2021-02-10]
21. Shaban-Nejad A, Michalowski M, Buckeridge DL. Health intelligence: how artificial intelligence transforms population and personalized health. NPJ Digit Med 2018;1:53 [FREE Full text] [doi: 10.1038/s41746-018-0058-9] [Medline: 31304332]
22. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]
23. Kürzinger ML, Schück S, Texier N, Abdellaoui R, Faviez C, Pouget J, et al. Web-Based Signal Detection Using Medical Forums Data in France: Comparative Analysis. J Med Internet Res 2018 Nov 20;20(11):e10466 [FREE Full text] [doi: 10.2196/10466] [Medline: 30459145]
24. Najork M. Web crawler architecture. In: Liu L, Özsu MT, editors. Encyclopedia of Database Systems. New York, NY: Springer; Sep 01, 2009:3462-3465.
25. Baby Center. URL: https://www.babycenter.com/pregnancy/week-by-week [accessed 2022-01-11]
26. Fleischer AB. Atopic dermatitis: the relationship to temperature and seasonality in the United States. Int J Dermatol 2019 Apr 11;58(4):465-471. [doi: 10.1111/ijd.14289] [Medline: 30417330]
27. Dhar S, Kanwar A. Epidemiology and clinical pattern of atopic dermatitis in a North Indian pediatric population. Pediatr Dermatol 1998 Sep;15(5):347-351. [doi: 10.1046/j.1525-1470.1998.1998015347.x] [Medline: 9796582]
28. Rollero C, Daniele A, Tartaglia S. Do men post and women view? The role of gender, personality and emotions in online social activity. Cyberpsychology 2019 Feb 21;13(1):1-15. [doi: 10.5817/cp2019-1-1]
29. Gender-Specific Behaviors on Social Media and What They Mean for Online Communications. Social Media Today. 2016 Nov 6. URL: https://www.socialmediatoday.com/social-networks/gender-specific-behaviors-social-media-and-what-they-mean-online-communications [accessed 2021-02-17]
30. Zuberbier T, Orlow SJ, Paller AS, Taïeb A, Allen R, Hernanz-Hermosa JM, et al. Patient perspectives on the management of atopic dermatitis. J Allergy Clin Immunol 2006 Jul;118(1):226-232. [doi: 10.1016/j.jaci.2006.02.031] [Medline: 16815160]
31. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
32. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of Health Misinformation on Social Media: Systematic Review. J Med Internet Res 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
33. Pulido CM, Ruiz-Eugenio L, Redondo-Sama G, Villarejo-Carballido B. A New Application of Social Impact in Social Media for Overcoming Fake News in Health. Int J Environ Res Public Health 2020 Apr 03;17(7):2430. [doi: 10.3390/ijerph17072430] [Medline: 32260048]

34. Pariser DM, Simpson EL, Gadkari A, Bieber T, Margolis DJ, Brown M, et al. Evaluating patient-perceived control of atopic dermatitis: design, validation, and scoring of the Atopic Dermatitis Control Tool (ADCT). Curr Med Res Opin 2020 Mar 12;36(3):367-376. [doi: 10.1080/03007995.2019.1699516] [Medline: 31778083]

Abbreviations

AD: atopic dermatitis
ADCT: Atopic Dermatitis Control Tool
AI: artificial intelligence
BTM: Biterm Topic Model
HRQoL: health-related quality of life
ISOLATE: International Study on Life with Atopic Eczema
QoL: quality of life
TF-IDF: term-frequency inverse document frequency

Edited by A Mavragani; submitted 11.06.21; peer-reviewed by M Janodia, K Rolls; comments to author 08.09.21; revised version received 04.10.21; accepted 30.11.21; published 28.01.22.

Please cite as:
Voillot P, Riche B, Portafax M, Foulquié P, Gedik A, Barbarot S, Misery L, Héas S, Mebarki A, Texier N, Schück S
Social Media Platforms Listening Study on Atopic Dermatitis: Quantitative and Qualitative Findings
J Med Internet Res 2022;24(1):e31140
URL: https://www.jmir.org/2022/1/e31140
doi:10.2196/31140
PMID:35089160

Original Paper

Heterogeneity of Prevalence of Social Media Addiction Across Multiple Classification Schemes: Latent Profile Analysis

Cecilia Cheng[1], PhD; Omid V Ebrahimi[2,3], MSc; Jeremy W Luk[4], PhD

[1]Department of Psychology, The University of Hong Kong, Pokfulam, Hong Kong

[2]Department of Psychology, University of Oslo, Oslo, Norway

[3]Research Institute, Modum Bad Psychiatric Hospital, Vikersund, Norway

[4]National Institute on Alcohol Abuse and Alcoholism, Bethesda, MD, United States

Corresponding Author:
Cecilia Cheng, PhD
Social and Health Psychology Lab
Department of Psychology
The University of Hong Kong
Pokfulam
China (Hong Kong)
Phone: 852 39174224
Email: ceci-cheng@hku.hk

Abstract

Background: As social media is a major channel of interpersonal communication in the digital age, social media addiction has emerged as a novel mental health issue that has raised considerable concerns among researchers, health professionals, policy makers, mass media, and the general public.

Objective: The aim of this study is to examine the prevalence of social media addiction derived from 4 major classification schemes (strict monothetic, strict polythetic, monothetic, and polythetic), with latent profiles embedded in the empirical data adopted as the benchmark for comparison. The extent of matching between the classification of each scheme and the actual data pattern was evaluated using sensitivity and specificity analyses. The associations between social media addiction and 2 comorbid mental health conditions—depression and anxiety—were investigated.

Methods: A cross-sectional web-based survey was conducted, and the replicability of findings was assessed in 2 independent samples comprising 573 adults from the United Kingdom (261/573, 45.6% men; mean age 43.62 years, SD 12.24 years) and 474 adults from the United States (224/474, 47.4% men; mean age 44.67 years, SD 12.99 years). The demographic characteristics of both samples were similar to those of their respective populations.

Results: The prevalence estimates of social media addiction varied across the classification schemes, ranging from 1% to 15% for the UK sample and 0% to 11% for the US sample. The latent profile analysis identified 3 latent groups for both samples: low-risk, at-risk, and high-risk. The sensitivity, specificity, and negative predictive values were high (83%-100%) for all classification schemes, except for the relatively lower sensitivity (73%-74%) for the polythetic scheme. However, the polythetic scheme had high positive predictive values (88%-94%), whereas such values were low (2%-43%) for the other 3 classification schemes. The group membership yielded by the polythetic scheme was largely consistent (95%-96%) with that of the benchmark.

Conclusions: Among the classification schemes, the polythetic scheme is more well-balanced across all 4 indices.

(J Med Internet Res 2022;24(1):e27000) doi:10.2196/27000

KEYWORDS

behavioral addiction; compulsive social media use; information technology addiction; mental health; psychological assessment; sensitivity; social network site; social networking; well-being

Introduction

Background Context

The internet has turned the world into an interconnected, global village in which information and problems alike spread swiftly across countries. Apart from face-to-face interactions, social media has emerged as a major channel of interpersonal communication in the cyber age. Social media use is beneficial in many aspects. For instance, social media allows people to maintain contact with existing friends or family members who live apart [1,2], and individuals who use social media to connect with their social network members tend to experience greater levels of subjective well-being [3,4]. In the extended period of social distancing during the COVID-19 pandemic, social media use is found to be associated with both social and physical well-being [5,6]. For people who seek mental health professional services, the interventions delivered through social media are deemed more accessible and engaging than face-to-face interventions [7].

Despite these psychological benefits, the misuse of social media can incur considerable psychological costs, especially for individuals who use social media as a refuge to evade unpleasant feelings or real-life problems. According to the model of compensatory internet use [8], social media use may improve these individuals' psychological condition in the short run, but such short-term benefits may also strengthen their dependence on social media, leading to continuous excessive use of social media and the development of social media addiction [9]. Moreover, social media addiction is motivated by the need to fulfill fundamental psychological needs that cannot be gratified in the real world, such as the need to belong [10-13].

Social Media Addiction and Its Assessment

Social media addiction is a type of behavioral addiction that is broadly defined as compulsive engagement in social media platforms that significantly disrupts the users' functioning in important life domains, such as interpersonal relations, work or study performance, and physical health [14,15]. According to the components model of behavioral addiction [16], social media addiction is conceptualized as a set of symptoms pertaining to six types of problematic behavior: (1) salience, which refers to the dominance of social media activities in one's thoughts and daily life; (2) tolerance, which refers to the tendency of spending an increasing amount of time using social media to attain the same amount of pleasure; (3) mood modification, which refers to the use of social media to avoid or mitigate unpleasant emotions experienced in real-life events; (4) relapse, which refers to the failure of curbing excessive social media use after attempts of abstinence or control; (5) withdrawal, which refers to psychological distress experienced when one cannot get access to social media; and (6) conflict, which refers to the adverse impact on one's job or studies because of problematic social media use.

Although social media addiction is not currently a diagnosable condition, researchers have constructed measures of social media addiction based on the diagnostic criteria for other behavioral addictions such as gambling disorder [16,17]. The most popular ones include the Facebook Intrusion Questionnaire [18] and the Bergen Social Media Addiction Scale (BSMAS) [19]. The former measure assesses addiction specific to a single social media platform, whereas the latter assesses addiction to social media in general. Instead of creating new assessment tools, another group of researchers have modified the current validated scales of internet addiction. The modification typically involves altering the context from *Internet* to *Facebook* or *social media*. For instance, the items of the Problematic Facebook Use Scale [20] were adapted from those of the Generalized Problematic Internet Use Scale [21].

Classification Schemes for Social Media Addiction

Validated measures of social media addiction have been widely used as screening tools for distinguishing individuals with and without the problem [22,23]. Both monothetic and polythetic formats have been adopted to yield prevalence estimates and to screen cases [24,25], because these schemes have long been used in case classifications of psychiatric disorders in the Diagnostic and Statistical Manual of Mental Disorders (DSM) [17,26,27]. The *classical* monothetic classification is generally regarded as more conservative because a positive diagnosis requires the endorsement of all the listed criteria [28]. For polythetic classification, however, no single criterion is required to make a diagnosis. The polythetic classification is more liberal than the monothetic classification because a positive diagnosis requires the endorsement of more than half of the listed criteria rather than all; therefore, individuals with the same classification may have different clinical presentations. Polythetic classification is commonly used in a variety of clinical diagnoses, including gambling disorder and substance abuse [17,27].

Most existing measures of social media addiction consist of items that are answered on a Likert-type scale ranging from 1 to 5 rather than dichotomous options. A usual practice for indicating the presence of a symptom involves the recoding of the 5-point ratings using the midpoint (ie, 3) as the cutoff such that a particular criterion is met for a score of 3 or above. Some researchers recently advocated stricter coding by setting a higher cutoff of 4 instead of the midpoint [29,30]. Taken together, a review of the literature reveals 4 commonly adopted classification schemes: monothetic, polythetic, strict monothetic, and strict polythetic. As these schemes vary in the extent of strictness in case classifications, different prevalence estimates are obtained, with higher prevalence yielded from more liberal classifications such as the polythetic scheme.

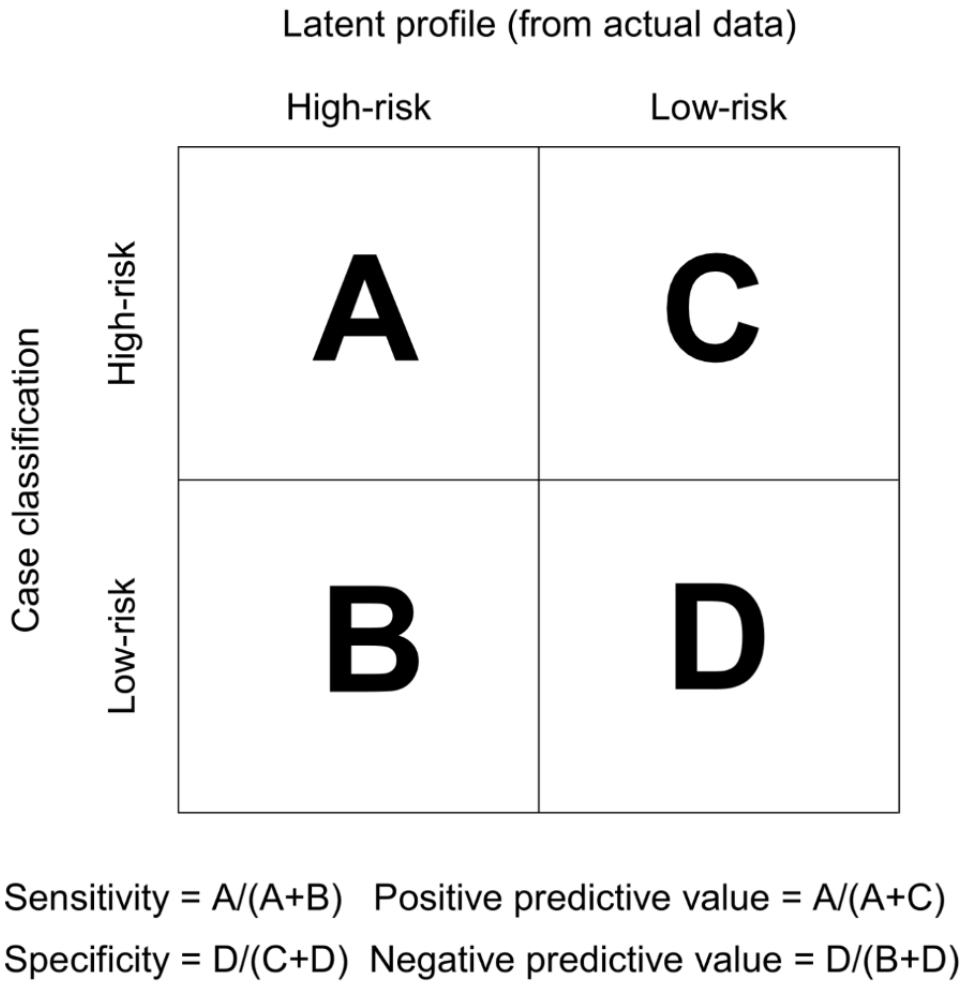
Previous studies on social media addiction have adopted either 1 or at most 2 of the classification schemes for deriving prevalence estimates, and the reported prevalence rates differ vastly across studies. As social media addiction is a global mental health concern, researchers worldwide have investigated the prevalence of this emergent problem [31-34]. The samples recruited in these prevalence studies vary considerably in their demographic characteristics, such as age and ethnicity, making between-study comparisons difficult. This study is the first to apply all 4 major classification schemes such that comparisons of the prevalence drawn from various schemes can be made. A total of 2 independent, demographically heterogeneous samples

were included to evaluate the extent of cross-sample replicability in the findings.

Evaluation of Classification Schemes for Social Media Addiction

Sensitivity and specificity analyses are widely used for the evaluation of classification schemes [35]. This method seeks to test the performance of a classification in matching and predicting a diagnosis or outcome. As a *gold standard* for classifying social media addiction is currently unavailable, the latent profiles embedded in a data set are adopted as a benchmark to evaluate which existing classification scheme for social media addiction can provide the best fit to empirical data. Latent profile analysis (LPA) is a type of mixture modeling that can reveal hidden subgroups of individuals from observed data that share similar symptom profiles [36]. This person-centered statistical approach is especially appropriate for classifying disorders with heterogeneous symptoms, because highly consistent classifications can be obtained by precise distinctions among profiles and differences in characteristics among profiles [37,38]. LPA has been applied to social media addiction and can be used to estimate the proportion of the population with different risk levels of this disorder [39]. This study extends the literature by evaluating the extent to which the prevalence estimate derived from a particular classification approximates the latent profiles of symptoms actually endorsed by respondents, and the results are indicated by 4 indices: sensitivity, specificity, positive predictive value, and negative predictive value (Figure 1).

Figure 1. The 4 indicators of sensitivity and specificity analysis.



In this study, sensitivity refers to the proportion of individuals from the data-driven, latent high-risk group who are also classified by a particular scheme as high risk. Specificity refers to the proportion of individuals from the data-driven, latent low-risk group who are also classified by a particular scheme as low risk. Positive predictive value refers to the proportion of high-risk participants classified by a particular scheme who also belong to the data-driven, latent high-risk group. Negative predictive value refers to the proportion of low-risk participants classified by a particular scheme who also belong to the data-driven, latent low-risk group.

Social Media Addiction and Comorbid Mental Health Problems

To establish that social media addiction is not merely a normative behavioral pattern, it is important to evaluate whether it is associated with other comorbid mental health problems. In the literature on social media addiction, a frequently researched psychiatric condition is depression, which has been found to have positive associations with social media addiction. Specifically, individuals with depression tend to have a high risk of developing social media addiction [40,41]. Although

users generally have more pleasant experiences than unpleasant ones when engaging in social media, unpleasant experiences on social media (eg, cyberbullying and social comparison with friends or strangers) tend to compromise mental health [42]. These adverse mental health impacts are especially salient among individuals with depression owing to their ruminative response style [43,44].

Another frequently researched psychiatric condition is anxiety [10,45]. Some studies have shown that individuals with anxiety are prone to social media addiction because of their strong motivation to avoid face-to-face social interactions [46]. As individuals with social media addiction tend to spend more time on web-based than in-person social interactions, their prolonged social media use may erode social skills and promote greater fear of meeting people in real life [47]. Their anxiety may in turn aggravate their symptoms of social media addiction, as these individuals continue to perceive in-person interactions as a source of threat [46]. People who fear face-to-face communication experience a deficit in need for relatedness, which may arise from feelings of insecurity in daily life situations, and social media may thus be used as a compensational tool to gratify their relatedness needs [47]. Prolonged, excessive use of social media may lead to social media addiction among individuals with heightened anxiety [48].

Methods

Participant Recruitment and Procedures

The 2 independent samples for this study were recruited from Prolific because this web-based participant pool provides heterogeneous samples with diverse demographic characteristics. Moreover, the participants of this web-based survey platform were found to be more honest and naïve, and their data quality was better than that of members from other survey platforms [49]. Eligible participants were residents of the United Kingdom or the United States aged between 18 and 65 years and were users of at least one social networking site. To maintain data quality, only those who had an approval rating of ≥90% on the survey site were included.

The samples were recruited from the United Kingdom and the United States because members of both countries are the largest body of consumers of the English version of the BSMAS. These countries are appropriate for sample replication because they are highly similar in their internet penetration rates, socioeconomic development, and cultural values [50,51]. Recruitment was carried out from May 18, 2020, to May 24, 2020.

All participants completed a set of questionnaires that was constructed and launched through a web-based survey tool, Qualtrics (Qualtrics International Incorporation). The research protocol of this study followed those adopted in previous web-based surveys [29,52]. An advertisement was placed on the survey platform, and those interested were invited to sign up. All participants had to give their consent before the survey began. Upon completion of the survey, they were paid according to the standard rate set by the survey platform. The human research ethics committee of the university where the principal author is employed approved the research protocol before this study was implemented.

Measures

Social Media Addiction

The symptoms of social media addiction were measured by the BSMAS [10], which is by far the most widely used validated assessment tool for addictive use of social media in general. This scale comprises 6 items, each measuring a core symptom (ie, salience, tolerance, mood modification, loss of control, withdrawal, and conflict). Item wording was consistent with the diagnostic criteria for gambling disorder [17]. The respondents rated each item on a 5-point scale, ranging from 1 (*very rarely*) to 5 (*very often*). The item scores were aggregated to obtain a composite score, and the 4 major classification schemes outlined in the *Introduction* section (the *Classification Schemes for Social Media Addiction* subsection) were adopted to categorize respondents into groups: high- versus no-to-low-risk. Prevalence rate refers to the proportion of participants who were classified as having social media addiction to the entire sample, and the prevalence rates derived from various classification schemes were then compared. The BSMAS was a reliable measure in this study (Cronbach α=.88 and .86 for the UK and the US samples, respectively).

Depression

In this study, depression and anxiety were included as criterion variables owing to the high comorbidity between social media addiction and both of these mental health conditions [53,54]. Probable depression was assessed using the Center for Epidemiological Studies Depression Scale [55], which was constructed for use with general, nonpsychiatric populations. Respondents were instructed to rate each of the 20 items on a 4-point scale ranging from 0 (*rarely or none of the time*) to 3 (*most or all of the time*). A higher composite score indicated a higher level of depression. This measure is widely adopted as a screening tool for clinical depression, with a recommended cutoff score of 16. This threshold score was thus adopted to indicate probable depression. The measure exhibited excellent psychometric properties in screening for major depression in the general population as verified by the DSM [55]. The depression measure displayed internal consistency in this study (Cronbach α=.92 and .91 for the UK and the US samples, respectively).

Anxiety

Probable anxiety was measured by the state anxiety subscale of the State-Trait Anxiety Inventory Form Y1 [56], which was selected because this extensively validated measure has by far been the most popular screening tool for anxiety [57]. Respondents rated each of the 20 items on a 4-point scale ranging from 1 (*not at all*) to 4 (*very much so*). A higher composite score indicates a higher level of anxiety. According to the community adult norms stated in the manual [56], a cutoff score of 40 was used for screening. This measure was found to be reliable (Cronbach α=.96 and .96 for the UK and the US samples, respectively).

Data Analysis

LPA was conducted because this person-centered approach is currently widely applied for identifying latent groups with similar characteristics [58]. A total of 6 BSMAS items were included as indicator variables in this analysis. Multiple indices were checked to determine the model with the best fit of data [59]. Specifically, a number of models with different class solutions (*k* ranging from 2 to 5) were tested, and better data fit was indicated by lower values of 3 goodness-of-fit statistics: Bayesian Information Criterion, sample size adjusted Bayesian Information Criterion, and Akaike Information Criterion. Entropy was also examined to evaluate the precision of assigning latent group membership, with a value of ≥0.80, indicating precision of classification. A class solution with an entropy value of <0.80 was ruled out. In addition, the Lo-Mendell-Rubin likelihood ratio test and bootstrap likelihood ratio test were used to assist model selection, and a significant result (ie, $P<.05$) showed that a *k* class model improved a *k*–1 class model. If the results revealed that 2 or more models were adequate, model parsimony and interpretability were considered. After the model selection decision had been made, the profiles were plotted for each group using a line graph.

The latent profiles identified in the LPA were then mapped onto existing classification schemes using sensitivity and specificity analyses [60]. A total of 4 indices—sensitivity, specificity, positive predictive value, and negative predictive value—were examined (the *Evaluation of Classification Schemes for Social Media Addiction* subsection and Figure 1). In addition, an index of overall consistency was reported to indicate the percentage of overlap in group membership between the latent profiles and high-versus-low-risk groups classified by a particular scheme.

Risk ratio was used to make two types of estimation: (1) the individual contribution of demographic variables (sex and age groups) to the risk for social media addiction and (2) the individual contribution of social media addiction prevalence to the risk for mental health problems. All statistical analyses were conducted using SPSS version 26 (IBM), except for LPA, which was conducted using MPlus version 8.5 (Muthén and Muthén).

Funding and Ethical Considerations

This study was funded in part by the General Research Fund administered by the Research Grants Council of Hong Kong in January 2020 (grant 17400714). The research protocol was reviewed and approved by the human research ethics committee of the University of Hong Kong before data collection (approval number: EA2002033**;** approval date: March 4, 2020). All study procedures were performed in accordance with the ethical principles of the Declaration of Helsinki. All participants were required to provide informed consent before completing the survey.

Results

Sample Characteristics

The UK sample comprised 573 adults, whereas the US sample comprised 474 adults. The sample size of each country met the requirements for conducting covariance modeling [59] and sensitivity and specificity analyses [35]. The average age of the participants in the UK sample was 43.62 years (SD 12.24 years) and that of the participants in the US sample was 44.67 years (SD 12.99 years). Table 1 presents the sex and age distribution of the participants from both samples as well as those of the UK and the US populations. As shown in Table 1, sample distributions of these major demographic variables were comparable to those of their respective populations.

Table 1. Sex and age distribution of the 2 samples compared with that of their own population (N=1047).

Parameters	The United Kingdom, n (%)		The United States, n (%)	
	Sample (n=573)	Population in 2020	Sample (n=474)	Population in 2020
Sex				
Male	261 (45.6)	33,821[a] (49.8)	224 (47.4)	165,899[a] (50.1)
Female	312 (54.4)	34,065 (50.2)	250 (52.6)	165,104 (49.9)
Age group (years)				
18-34	192 (33.5)	22,775 (33.5)	169 (35.6)	118,036 (35.6)
35-49	191 (33.3)	22,654 (33.4)	152 (32.1)	105,788 (32)
50-65	190 (33.2)	22,457 (33.1)	153 (32.3)	107,179 (32.4)

[a]Figures in these columns are expressed in thousands.

Case Classification With LPA

Table 2 presents the results of LPA that were tested and compared among the 4 models derived from different classification schemes. For the UK sample, the 3-class model was selected because the Lo-Mendell-Rubin likelihood ratio test showed that this model had better data fit than the 2-class model, but the degree of data fit was highly similar for the 3-class, 4-class, and 5-class models. Hence, the 3-class model was chosen for parsimonious considerations.

Table 2. Summary of latent profile analysis comparing the various models (N=1047).

Characteristics	Model comparison			
	2-class model	3-class model	4-class model	5-class model
The UK Sample (n=573)				
BIC[a]	8748.65	8196.68	8063.43	7882.23
SSABIC[b]	8688.33	8114.14	7958.67	7755.25
AIC[c]	8665.99	8083.56	7919.85	7708.20
Entropy	0.94	0.90	0 90	0.93
LMR-LRT[d], *P* value	.004	.001	.32	.08
BLRT[e], *P* value	<.001	<.001	<.001	<.001
The US Sample (n=474)				
BIC	6755.99	6342.76	5946.03	5855.20
SSABIC	6695.69	6260.24	5841.29	5728.24
AIC	6676.93	6234.56	5808.71	5688.75
Entropy	0.93	0.91	0.93	0.93
LMR-LRT, *P* value	<.001	.02	.005	.09
BLRT, *P* value	<.001	<.001	<.001	<.001

[a]BIC: Bayesian Information Criteria.

[b]SSABIC: sample size adjusted Bayesian Information Criterion.

[c]AIC: Akaike Information Criterion.

[d]LMR-LRT: Lo-Mendell-Rubin likelihood ratio test.

[e]BLRT: bootstrap likelihood ratio test.

The latent profiles of the participants from the United Kingdom are shown in Figure 2. As shown in this figure, the 3 latent groups differed in terms of symptom severity. Specifically, more than half of the participants were assigned to the first group (315/573, 55%) characterized by low mean item scores across all 6 criteria of social media addiction (<1.34), and this group was labeled as *low-risk*. The profile pattern of the second group (190/573, 33.1%) was relatively more complex than that of the first group. The participants from the second group were more likely to endorse the salience, tolerance, and mood modification criteria, with item mean scores clustered around the midpoint (range 2.62-2.97), but not the remaining 3 criteria (mean item scores<2.07). The second group was labeled as *at-risk* because the mean item scores for half of the criteria approached the cutoff (ie, midpoint) for both the monothetic and polythetic schemes. The third group (68/573, 11.9%) had high mean item scores for all the criteria that were above the cutoff for the monothetic and polythetic schemes (>3.18), and this group was labeled as *high-risk*.

For the US sample, both the 3-class and the 4-class models demonstrated good data fit, but the former model was chosen because its grouping of participants was more interpretable than that of the latter. The 3 latent groups are shown in Figure 3. As revealed in this figure, the profiles of the first 2 groups were highly similar for the UK and the US samples. Specifically, the low-risk group of the US sample (295/474, 62.2%) had low mean item scores across all 6 criteria (<1.45), and the at-risk group (136/474, 28.7%) tended to endorse the same 3 criteria (salience, tolerance, and mood modification; mean item scores ranged from 2.78 to 3.08) but not the other 3 (mean item scores<1.89).

Figure 2. Latent profiles for the 3-class solution for the UK samples.

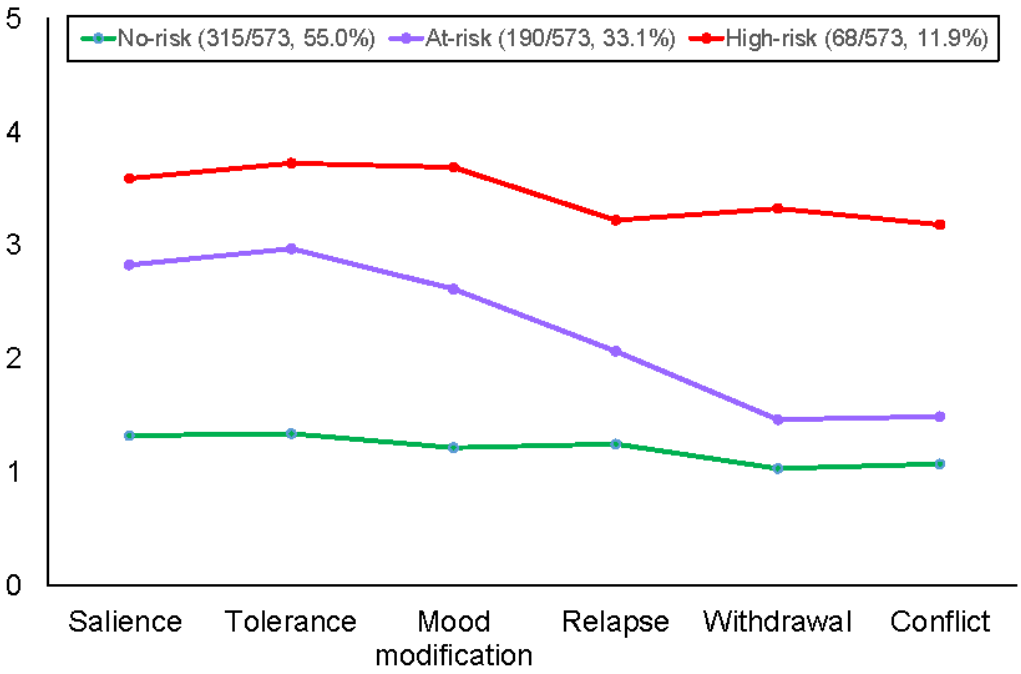
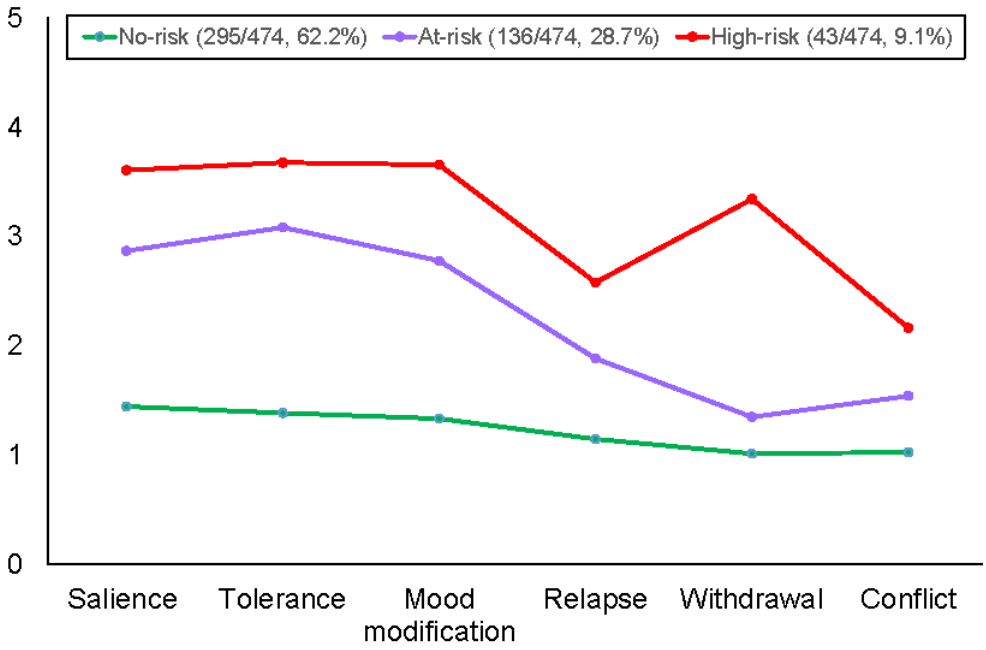


Figure 3. Latent profiles for the 3-class solution for the US samples.



Although the profiles of the low-risk and at-risk groups of participants from the United States were comparable with those of the same 2 groups of the participants from the United Kingdom, the profiles of the third group were different for the 2 samples. Instead of consistently having all scores above the midpoint across all 6 items as their UK counterparts did, the third group of the US sample (43/474, 9.1%) had mean item scores above the midpoint only for two-thirds of the criteria (ie, salience, tolerance, mood modification, and withdrawal; mean item scores>3.34). The third group of the US sample was also labeled as *high-risk* owing to their members' endorsement of 4 out of 6 criteria that was consistent with the polythetic scheme.

Ad hoc Bonferroni tests were conducted among the 3 latent groups for each of the 6 criteria. The results consistently showed that all the criterion scores of the high-risk group were significantly higher than those of the at-risk group, whose criterion scores were in turn significantly higher than those of the low-risk group (P<.001). These results provided evidence for the distinctness of the profiles of the 3 latent groups.

Prevalence of Social Media Addiction

As both of the samples were heterogeneous in terms of demographic characteristics, risk ratios were computed to identify sex and age differences. The results revealed considerable age differences in the prevalence of social media addiction for both samples (Table 3). For the UK sample, younger adults aged between 18 and 34 years were at a 4-fold higher risk of social media addiction than their older counterparts aged between 50 and 65 years, and this pattern of findings was consistent across the various classification schemes. For the US sample, the risk of social media addiction was 2 to 4 times higher in the group of younger adults compared with the group of older adults. However, sex differences in the prevalence of social media addiction were not prominent for both samples, except female participants from the US sample were found to have higher risks of social media addiction than their male counterparts when the strict polythetic scheme was adopted for classification.

Table 3. Prevalence of social media addiction by major classification scheme, sex, and age group for the 2 samples (N=1047).

Parameters	Classification scheme									
	Strict monothetic, n (%)	Strict monothetic, RR[a]	Strict polythetic, n (%)	Strict polythetic, RR	Monothetic, n (%)	Monothetic, RR	Polythetic, n (%)	Polythetic, RR	Latent profile[b], n (%)	Latent profile[b], RR
The UK Sample (n=573)										
Total	4 (0.7)	—[c]	23 (4)	—	29 (5.1)	—	86 (15)	—	68 (11.9)	—
Sex										
Male	7 (1.2)	N/A[d]	28 (4.9)	1.56	35 (6.1)	1.58	101 (17.6)	1.53	72 (12.6)	1.18
Female	0 (0)	Ref[e]	18 (3.1)	Ref	22 (3.8)	Ref	66 (11.5)	Ref	62 (10.8)	Ref
Age group (years)										
18-34	12 (2.1)	4.03	48 (8.4)	4.09	60 (10.5)	4.09	144 (25.1)	3.78	132 (23)	6.20
35-49	0 (0)	N/A	17 (3)	1.45	22 (3.8)	1.49	90 (15.7)	2.36	63 (11)	2.41
50-65	3 (0.5)	Ref	12 (2.1)	Ref	15 (2.6)	Ref	38 (6.6)	Ref	26 (5)	Ref
The US sample (n=474)										
Total	1 (0.2)	—	12 (2.5)	—	12 (2.5)	—	52 (11)	—	43 (9.1)	—
Sex										
Male	2 (0.4)	N/A	21 (4.4)	4.38	13 (2.7)	1.23	64 (13.5)	1.66	47 (9.9)	1.25
Female	0 (0)	Ref	4 (0.8)	Ref	11 (2.3)	Ref	39 (8.2)	Ref	39 (8)	Ref
Age group (years)										
18-34	0 (0)	N/A	16 (3.4)	2.24	20 (4.2)	4.19	72 (15.2)	2.16	60 (12.7)	2.74
35-49	3 (0.6)	N/A	15 (3.2)	2.09	15 (3.2)	3.13	60 (12.7)	1.79	54 (11.4)	1.13
50-65	0 (0)	Ref	7 (1.5)	Ref	5 (1.1)	Ref	34 (7.2)	Ref	24 (5.1)	Ref

[a]RR: risk ratio.

[b]The low-risk and at-risk groups were coded as 0, and the high-risk group was coded as 1.

[c]No reference group.

[d]N/A: not applicable (cannot be computed).

[e]Ref: reference group.

Sensitivity and Specificity Analysis

The sensitivity and specificity analyses are summarized in Table 4. The various classification schemes generally had a high sensitivity (83%-100%) for both the UK and the US samples, with the exception of the polythetic scheme that had somewhat lower sensitivity (74% and 73% for the UK and the US samples, respectively). The specificity and negative predictive value were also high across the various schemes (>88% and >95%). Moreover, there were considerable consistencies or overlaps (>88%) in group membership between the latent profiles and the high-versus-low-risk groups classified by all the schemes, indicating that the participants classified by the various schemes were largely consistent with the latent groups embedded in the data when their sensitivity, specificity, and negative predictive value were examined.

Table 4. Sensitivity and specificity analyses of various classification schemes with the latent group as the benchmark (N=1047).

Sample and classification scheme	Indicator of sensitivity and specificity analyses (%)				
	Sensitivity	Specificity	Positive predictive value	Negative predictive value	Overall consistency
The UK sample (n=573)					
Strict monothetic	100	89	6	100	89
Strict polythetic	100	92	34	100	92
Monothetic	100	93	43	100	93
Polythetic	74	99	94	96	95
The US sample (n=474)					
Strict monothetic	100	91	2	100	91
Strict polythetic	83	93	23	100	93
Monothetic	100	93	28	100	93
Polythetic	73	99	88	97	96

Despite these similarities among the schemes, it is important to note that the schemes differed vastly in their positive predictive value. Only the polythetic scheme had a high positive predictive value for both samples (>88%), whereas the other schemes had a low positive predictive value (<43%). These results indicated that only the polythetic scheme could identify high-risk participants that were largely consistent with the data, whereas all other schemes might fail to identify a considerable proportion of participants who were classified as high risk by the data-driven LPA.

In summary, the various classification schemes generally had good sensitivity and specificity, but their positive predictive value was low, indicating that those participants who were classified by these schemes as having high risks only represented a relatively small proportion of participants from the data-driven, latent high-risk group. Despite having somewhat lower sensitivity than other schemes, the polythetic scheme had a high positive predictive value, indicating that the membership of various groups derived from the polythetic scheme overlapped with the group membership identified by the data-driven latent profiles as much as 96% of the time. Taken together, the polythetic scheme yielded the best balance of sensitivity, specificity, positive predictive value, and negative predictive value among the major classification schemes, and may thus be optimal for classifying cases of social media addiction.

Risk Ratio of Mental Health Problems by Classification Scheme and Criteria

Table 5 presents the descriptive statistics of mental health problems for both low-to-at-risk and high-risk groups classified by various schemes, whereas Table 6 displays the risk ratios of various major mental health problems associated with the incidence of social media addiction at both the scale and item levels. In these analyses, the reference group referred to the low-to-at-risk group classified by various schemes of social media addiction.

Table 5. Descriptive statistics of mental health problems by case classification of major schemes for the 2 samples (N=1047).

Scheme and mental health problem	The UK sample (n=573)			The US sample (n=474)		
	High-risk group, mean (SD)	Low-to-at-risk group, mean (SD)	*P* value[a]	High-risk group, mean (SD)	Low-to-at-risk group, mean (SD)	*P* value
Strict monothetic						
Depression	41.25 (9.64)	16.87 (11.08)	<.001	N/A[b]	N/A	N/A
Anxiety	64.25 (12.61)	40.07 (13.11)	<.001	N/A	N/A	N/A
Strict polythetic						
Depression	32.39 (11.26)	16.40 (10.79)	<.001	26.08 (8.62)	16.11 (10.89)	.002
Anxiety	56.13 (13.88)	39.56 (12.80)	<.001	48.25 (14.52)	38.39 (13.12)	.01
Monothetic						
Depression	27.03 (13.30)	16.51 (10.88)	<.001	26.00 (11.08)	16.12 (10.83)	.002
Anxiety	52.90 (14.20)	39.55 (12.86)	<.001	52.00 (11.95)	38.32 (13.11)	.001
Polythetic						
Depression	26.02 (12.67)	15.47 (10.21)	<.001	24.65 (10.38)	15.36 (10.58)	<.001
Anxiety	48.35 (13.94)	38.76 (12.59)	<.001	45.25 (13.09)	37.83 (13.04)	<.001
Latent profile[c]						
Depression	26.59 (12.70)	15.75 (10.39)	<.001	24.00 (11.02)	15.62 (10.65)	<.001
Anxiety	50.10 (14.01)	38.87 (12.56)	<.001	45.98 (13.07)	37.91 (13.04)	<.001

[a]*P* value indicates the significance level of an independent sample *t* test (1-tailed) for each mental health problem of a sample.

[b]N/A: not applicable (cannot be computed).

[c]The low-risk and at-risk groups were coded as 0, and the high-risk group was coded as 1.

Table 6. Risk ratio of mental health problems by classification scheme and criteria for social media addiction (N=1047).

Parameters	Mental health problems in the UK sample (n=573), RR[a]		Mental health problems in the US sample (n=474), RR	
	Probable depression	Probable anxiety	Probable depression	Probable anxiety
Classification scheme[b]				
Strict monothetic	N/A[c]	N/A	N/A	N/A
Strict polythetic	5.99	6.17	6.59	14.81
Monothetic	2.77	8.48	3.92	2.90
Polythetic	4.46	3.27	6.21	2.65
Latent profile	4.17	3.85	4.60	2.79
Criteria[d]				
Salience	1.97	2.18	2.51	1.68
Tolerance	2.58	2.16	3.35	2.08
Mood modification	4.34	3.54	6.33	2.99
Loss of control	1.57	1.38	2.62	2.29
Withdrawal	5.99	3.52	7.29	8.95
Conflict	7.02	2.79	3.88	4.33
Criteria[e]				
Salience	1.44	1.67	2.19	1.66
Tolerance	1.29	1.63	2.82	1.81
Mood modification	3.93	2.83	5.22	3.20
Loss of control	2.37	1.87	3.06	3.09
Withdrawal	3.17	3.54	4.60	2.79
Conflict	4.08	3.79	3.30	2.65

[a]RR: risk ratio, with the low-risk and at-risk groups coded as 0, and the high-risk group was coded as 1.

[b]This analysis was conducted at the scale level.

[c]N/A: not applicable (cannot be computed).

[d]This analysis was conducted at the item level with a cutoff score of 4.

[e]This analysis was conducted at the item level with a cutoff score of 3.

For the UK sample, the risk of probable depression or anxiety was about 3 to 8 times higher in the high-risk group by various schemes than in the low-to-at-risk group. Similarly, for the US sample, the risk of having any of these mental health problems was about 3 to 15 times higher for the high-risk (vs no-to-low-risk) group identified by various schemes.

As the profiles of the latent groups revealed some interesting patterns across the 6 criteria for social media addiction, additional analyses were conducted at the item (criterion) level. Specifically, each of the BSMAS items was dummy coded according to the cutoff adopted in the strict monothetic and strict polythetic schemes (ie, 4 out of a 5-point scale). As shown in the middle panel of Table 6, the risk of probable depression or anxiety was about 3 to 9 times higher for the mood modification, withdrawal, and conflict criteria for both the UK and the US samples when a high cutoff of 4 was applied.

The BSMAS items were also dummy coded according to the cutoff adopted in the monothetic and the polythetic schemes (ie, 3 out of a 5-point scale), and the results are summarized in the lower panel of Table 6. Similar to the findings derived from a higher cutoff of 4, the risk of probable depression or anxiety was about 3 to 5 times higher for the mood modification, withdrawal, and conflict criteria for both the UK and the US samples.

Discussion

Principal Findings

Social media addiction has emerged as a prevalent problem of public concern in the modern cyber era, and this emergent problem has been examined in the context of the COVID-19 pandemic. Although the prevalence rates of social media addiction for both samples obtained in this period are comparable with those derived from the same countries before the pandemic [61], the psychiatric problems reported by both samples are more prevalent than those reported in previous studies [62]. These findings indicate that the residents of the United Kingdom and the United States could be emotionally overwhelmed by enormous stressors elicited during the early

phase of the pandemic [62,63], but the prevalence of social media addiction remained largely stable at that stage.

This study is the first to adopt a nuanced analysis of the various schemes that have been widely used for case classification purposes. Our findings indicate that the application of diverse schemes yields varied prevalence estimates, and testing these schemes against the data-driven, latent profile analytic approach thus generates valuable information that unveils the relative performance of different schemes in case classification.

In our study, three latent profiles are found to be embedded in the data: low-risk, at-risk, and high-risk. It is noteworthy that the shape of the symptom profile of the at-risk group was distinct from that of the other 2 groups. For the at-risk group, the endorsement of half of the criteria (ie, salience, tolerance, and mood modification) tended to be more similar to the endorsement of those by the high-risk group, whereas the endorsement of the other half (ie, loss of control, withdrawal, and conflict) by the at-risk group tended to be more similar to the endorsement of those by the low-risk group. These findings show that a dichotomous (low-risk vs high-risk) classification scheme may be inadequate. Instead, a tripartite classification scheme may be more appropriate for capturing respondents' distinct symptom characteristics of social media addiction, especially for the at-risk group whose symptom profile is similar to the low-risk and high-risk groups in certain criteria but not others.

In general, most classification schemes used in existing studies have good sensitivity and specificity when compared with the latent groups identified in our latent profile analyses. However, a low positive predictive value is found in most of the schemes, with the exception of the polythetic scheme. The present findings indicate that both strict monothetic and monothetic schemes have perfect sensitivity (100%) but at the cost of low positive predictive values (2%-43%). These 2 schemes may be too conservative, such that a large proportion of individuals with probable social media addiction are excluded from further assessment or follow-ups. In contrast, despite having slightly lower sensitivity than the other schemes, the polythetic scheme appears to have a more balanced performance across all psychometric indicators.

Research and Practical Implications

These new findings have research and practical implications. Specifically, the study delineates the strengths and weaknesses of each classification scheme, thus guiding decisions for selecting an optimal scheme for making population-level estimates of social media addiction. The findings indicate that the various major classification schemes have a large degree of heterogeneity, both in terms of estimating the prevalence rates of social media addiction as well as screening and detecting cases of social media addiction. With regard to prevalence estimates, researchers utilizing strict classification schemes may tend to obtain very low prevalence rates of social media addiction. As shown in this study, such discrepancies can be as large as 10% to 15% (eg, using polythetic vs strict monothetic schemes). It is noteworthy that the polythetic scheme was found to have relatively high consistency with the benchmark data of the 2 independent samples, with discrepancies of only 2% to 3% (11% vs 9% in the United States; 15% vs 12% in the United Kingdom). Thus, the present findings provide some empirical evidence that polythetic classification may best reflect classifications identified using a data-driven approach and is optimal when the goal is to identify a broad group of individuals at risk for social media addiction.

Furthermore, the assessment of social media addiction has become an issue of growing importance for mental health professionals, with some countries formulating public health policies and providing extensive training to guide professionals in distinguishing clients with varying severity levels of the problem [64]. As precise case classification is crucial for appropriate referral to tailored intervention, a better understanding of how well a screening instrument function using different scoring methods becomes important. In evaluating screening tools, the sensitivity and specificity of classification schemes have traditionally been emphasized. However, basing classification schemes solely on these 2 indices is insufficient for making clinical decisions at the individual level and may even be misleading in some cases [65]. This is because both indicators of sensitivity and specificity are population-based indices that represent properties of the screening test in itself [66], but the 2 predictive values further include sample-relevant information, such as the base rate of a given problem area [67]. Therefore, the use of multiple indices, including positive and negative predictive values, is a notable strength of our analyses, providing valuable data that could inform the scoring and interpretation of the BSMAS within the context of clinical assessment [65,68].

The BSMAS is a brief measure that has been commonly adopted to assess symptoms of social media addiction, although our results highlight several considerations relevant to clinicians' choice of classification schemes. Despite demonstrating high sensitivity and specificity, the use of the strict monothetic, strict polythetic, and monothetic schemes are found to be more prone to missing members of the high-risk group derived from the actual data (ie, positive predictive values ranging from 2% to 43%) as compared with the polythetic scheme (ie, positive predictive values of 88% and 94%). These findings demonstrate that the polythetic scheme has superior positive predictive rates of social media addiction compared with the other schemes, indicating the greater utility of the polythetic scheme in detecting individuals who are at high risk for social media addiction.

The polythetic scheme as compared with the monothetic scheme is more consistent with the current theoretical approaches to psychological functioning and mental health, as the polythetic scheme emphasizes a prototypical perspective (ie, requiring the presence of many rather than all symptoms) rather than a classical, monothetic perspective (ie, requiring the presence of all symptoms) to mental health [69]. Characterized by a prototypical approach, the polythetic scheme does not necessitate the presence of all symptoms within a problem area, representing an approach that is more attuned toward individual variability and heterogeneity, both of which are frequently observed in clinical and research settings. Similar to many other problem domains, the present findings indicate that social media addiction may be conceptualized as a continuum ranging from no to high risks, with a considerable proportion of at-risk

individuals clustered somewhere between the 2 extremes. Hence, the polythetic classification scheme is more similar to current assessment approaches to mental health problems, such as the DSM-5 and the 11th version of the International Statistical Classification of Diseases and Related Health Problems (ICD-11) [27,70]

Although information technology addiction (eg, social media addiction and gaming disorder) is currently not a diagnosable condition in the DSM-5 [71], a recent expert review has outlined meta-level criteria for determining whether different behavioral addictions may warrant the designation of *other specified disorders due to addictive behaviors* under the ICD-11 [72]. To determine whether social media addiction is present based on specific symptoms, this study fills the knowledge gap by revealing 3 symptoms of social media addiction with high-risk ratios with comorbid problems of depression and anxiety: mood modification, withdrawal, and conflict (Table 6). The replicable findings across the 2 independent samples suggest that these symptoms are crucial for the assessment of social media addiction. More importantly, these findings echo the classifications of a related problem of gaming disorder and other behavioral disorders in the ICD-11, in which conflict and adverse consequences in significant life domains are essential for a diagnosis.

This study further reveals that individuals identified with social media addiction have a high risk of probable depression and anxiety. Primarily, these findings indicate that the classification schemes of social media addiction may serve as a risk indicator in the screening process for potentially detecting comorbid problems with depression and anxiety. More importantly, several specific symptoms of social media addiction tend to have stronger connections with depression and anxiety, including mood modification, withdrawal, and conflict. Our item-level analyses identified several associations that warrant future research and attention by clinicians. First, the association between the specific symptoms of conflict and depression and anxiety corroborates those that have been empirically unveiled in a variety of contexts [73,74], highlighting conflict arising from social media addiction as a potential pathway linked with both mental health conditions. Clinicians may need to evaluate how conflict arising from social media addiction is related to depressive symptoms, such as the feelings of letting oneself and their significant others down, feelings of worthlessness, or anxiety symptoms associated with the monitoring of worry and rumination thoughts that are attributable to addictive use of social media [75,76]. Second, mood modification through addictive use of social media to avoid facing real-life problems is associated with a high risk of probable depression and anxiety. This finding may reflect individuals who use social media as a refuge to evade real-life challenges, duties, or responsibilities. If this is the case, clinicians may need to address social media addiction as a maladaptive, avoidant coping response that likely serves as a maintaining factor for both depression and anxiety [77,78]. Finally, as withdrawal symptoms are associated with psychological distress, individuals who are at risk for social media addiction may benefit from developing alternative coping skills that can replace social media activities when experiencing withdrawal.

In addition to the use of the BSMAS as a diagnostic tool for identifying individuals with social media addiction for treatment referral, this screening tool can also serve as an effective tool for early screening of at-risk cases to prevent further development into social media addiction. As prevention is often more cost-effective than treatment [79], broad-based screening can be valuable when mental health intervention resources are available in the community to serve the identified at-risk individuals. Given that this study evaluates the performance of multiple classification schemes against the empirically-derived benchmark, the findings can provide useful guidance for health care professionals to select the scheme most appropriate for their intervention schemes.

Research Limitations and Directions

Before concluding, caution should be exercised. First, this study adopted a quantitative design that included only validated measures with structured close-ended questions. The present inquiry thus focused on the typical symptoms of social media addiction, anxiety, and depression. The widely adopted quantitative design should be supplemented with qualitative data collection methods, such as narrative interviews and focus groups, which can broaden the scope of inquiry by unveiling participants' unique experiences [80,81]. Thus, a mixed methods design is encouraged to combine quantitative and qualitative methods to gain a more comprehensive perspective on social media addiction.

Second, our study used a web-based survey method for data collection and is thus vulnerable to the shortcomings inherent in this type of method. Although the screening function of the web-based survey platform allows the recruitment of the present samples whose demographic profiles resemble those of their respective populations, it is important to reiterate that nonprobabilistic sampling method was used in participant recruitment. Participants signed up for the web-based survey voluntarily upon placing an advertisement on the website, and such self-selection could potentially elicit sample bias.

Third, our study adopted a symptom approach in the examination of social media addiction and its mental health implications. Validated measures assessing a standard set of symptoms of social media addiction and the 2 psychiatric problems were administered. It is noteworthy that we did not include any measures of time for overall internet use or social media engagement [82,83]; thus, the amount and pattern of social media use were not assessed. For criterion assessment, only anxiety and depression were measured because both are major psychiatric comorbidities of information technology addiction [84,85]. Apart from psychiatric problems, individuals with information technology addiction also experience disruptions in other life domains, such as interpersonal relations and job performance [86,87]. The scope of daily life dysfunction should be broadened by including a greater variety of life domains for a more comprehensive evaluation of daily life challenges experienced by individuals with social media addiction.

Fourth, it is noteworthy that this study was conducted during the COVID-19 pandemic. As the massive global transmission of this unknown virus was unprecedented, an avalanche of false

and misleading information was disseminated through social media. Frequent use of social media was associated with COVID-19 anxiety that disturbed sleep quality [88,89]. Such psychological responses may not be fully captured by the traditional measures of mental health. Future research may consider using measures that capture COVID-19–specific stressors and experiences, such as loss of family and friends because of COVID-19 or exposure to misinformation about the pandemic via social media use [90].

Finally, although efforts have been made to replicate the findings in 2 independent samples from English-speaking countries with high internet penetration rates [50], the present findings cannot be generalized to individuals from other countries or cultural backgrounds. This is particularly the case as multinational meta-analyses have revealed the prevalence of information technology addiction and its differential underlying psychological mechanisms across cultural regions [61,84]. Therefore, future research should be expanded to include more countries with varying levels of cultural individualism and internet penetration. Data derived from a myriad of countries with diverse backgrounds enable researchers to make cross-cultural comparisons at both the individual and country levels [91].

Conclusions

In conclusion, this study evaluated 4 major schemes widely adopted to classify cases of social media addiction. Using latent profiles identified from empirical data as a benchmark, the performance of the polythetic scheme is more well-balanced in attaining relatively high levels of sensitivity, specificity, positive predictive value, and negative predictive value compared with those of the other 3 schemes. Although these findings are largely replicable in the 2 independent samples, efforts should be made to expand the scope of inquiry in countries with diverse cultural backgrounds using a wider range of criterion variables through qualitative methods, thus enriching the discussion and informing future decisions about the potential inclusion of social media addiction in the future versions of the DSM or ICD.

Acknowledgments

This study was funded by the General Research Fund administered by the Research Grants Council of Hong Kong (grant 17400714). The authors would like to thank Sylvia Lam, Sophie Lau, Janice Leung, and Olivia Qiu for their research and clerical assistance.

Conflicts of Interest

None declared.

References

1. Kim H. Enacted social support on social media and subjective well-being. Int J Commun 2014;8:21.
2. Ellison N, Steinfield C, Lampe C. The benefits of Facebook "friends": social capital and college students' use of online social network sites. J Comput Mediat Commun 2007 Jul 1;12(4):1143-1168. [doi: 10.1111/j.1083-6101.2007.00367.x]
3. Cheng C, Wang H, Sigerson L, Chau C. Do the socially rich get richer? A nuanced perspective on social network site use and online social capital accrual. Psychol Bull 2019 Jul;145(7):734-764. [doi: 10.1037/bul0000198] [Medline: 31094537]
4. Rae JR, Lonborg SD. Do motivations for using Facebook moderate the association between Facebook use and psychological well-being? Front Psychol 2015 Jun 12;6:771 [FREE Full text] [doi: 10.3389/fpsyg.2015.00771] [Medline: 26124733]
5. Cheng C, Lau Y, Luk JW. Social capital-accrual, escape-from-self, and time-displacement effects of internet use during the COVID-19 stay-at-home period: prospective, quantitative survey study. J Med Internet Res 2020 Dec 24;22(12):e22740 [FREE Full text] [doi: 10.2196/22740] [Medline: 33320824]
6. Norbury A, Liu SH, Campaña-Montes JJ, Romero-Medrano L, Barrigón ML, Smith E, MEmind Study Group, et al. Social media and smartphone app use predicts maintenance of physical activity during Covid-19 enforced isolation in psychiatric outpatients. Mol Psychiatry 2021 Aug 14;26(8):3920-3930 [FREE Full text] [doi: 10.1038/s41380-020-00963-5] [Medline: 33318619]
7. Ridout B, Campbell A. The use of social networking sites in mental health interventions for young people: systematic review. J Med Internet Res 2018 Dec 18;20(12):e12244 [FREE Full text] [doi: 10.2196/12244] [Medline: 30563811]
8. Kardefelt-Winther D. A conceptual and methodological critique of internet addiction research: towards a model of compensatory internet use. Comput Human Behav 2014 Feb;31:351-354. [doi: 10.1016/j.chb.2013.10.059]
9. Ryan T, Chester A, Reece J, Xenos S. The uses and abuses of Facebook: a review of Facebook addiction. J Behav Addict 2014 Sep;3(3):133-148 [FREE Full text] [doi: 10.1556/JBA.3.2014.016] [Medline: 25317337]
10. Schou Andreassen C, Billieux J, Griffiths MD, Kuss DJ, Demetrovics Z, Mazzoni E, et al. The relationship between addictive use of social media and video games and symptoms of psychiatric disorders: a large-scale cross-sectional study. Psychol Addict Behav 2016 Mar;30(2):252-262. [doi: 10.1037/adb0000160] [Medline: 26999354]
11. Gil-Or O, Levi-Belz Y, Turel O. The "Facebook-self": characteristics and psychological predictors of false self-presentation on Facebook. Front Psychol 2015 Feb 17;6:99 [FREE Full text] [doi: 10.3389/fpsyg.2015.00099] [Medline: 25741299]
12. Kuss DJ, Griffiths MD. Online social networking and addiction--a review of the psychological literature. Int J Environ Res Public Health 2011 Sep;8(9):3528-3552 [FREE Full text] [doi: 10.3390/ijerph8093528] [Medline: 22016701]

13. Pelling EL, White KM. The theory of planned behavior applied to young people's use of social networking web sites. Cyberpsychol Behav 2009 Dec;12(6):755-759. [doi: 10.1089/cpb.2009.0109] [Medline: 19788377]
14. Andreassen CS. Online social network site addiction: a comprehensive review. Curr Addict Rep 2015 Jun 01;2(2):175-184. [doi: 10.1007/s40429-015-0056-9]
15. Kuss D, Griffiths M. Social networking sites and addiction: ten lessons learned. Int J Environ Res Public Health 2017 Mar 17;14(3):311 [FREE Full text] [doi: 10.3390/ijerph14030311] [Medline: 28304359]
16. Griffiths M. A 'components' model of addiction within a biopsychosocial framework. J Substance Use 2005 Jul 12;10(4):191-197. [doi: 10.1080/14659890500114359]
17. Guze S. Diagnostic and statistical manual of mental disorders, 4th edition (DSM-IV). AJP 1995 Aug;152(8):1228. [doi: 10.1176/ajp.152.8.1228]
18. Elphinston RA, Noller P. Time to face it! Facebook intrusion and the implications for romantic jealousy and relationship satisfaction. Cyberpsychol Behav Soc Netw 2011 Nov;14(11):631-635. [doi: 10.1089/cyber.2010.0318] [Medline: 21548798]
19. Andreassen CS, Pallesen S, Griffiths MD. The relationship between addictive use of social media, narcissism, and self-esteem: findings from a large national survey. Addict Behav 2017 Jan;64:287-293. [doi: 10.1016/j.addbeh.2016.03.006] [Medline: 27072491]
20. Marino C, Vieno A, Altoè G, Spada MM. Factorial validity of the Problematic Facebook Use Scale for adolescents and young adults. J Behav Addict 2017 Mar 01;6(1):5-10 [FREE Full text] [doi: 10.1556/2006.6.2017.004] [Medline: 28198639]
21. Caplan S. Theory and measurement of generalized problematic internet use: a two-step approach. Comput Human Behav 2010 Sep;26(5):1089-1097. [doi: 10.1016/j.chb.2010.03.012]
22. van den Eijnden RJ, Lemmens JS, Valkenburg PM. The social media disorder scale. Comput Human Behav 2016 Aug;61:478-487. [doi: 10.1016/j.chb.2016.03.038]
23. Sahin C. Social media addiction scale-student form: the reliability and validity study. Turk Online J Educ Technol 2018 Jan;17(1):169-182.
24. Andreassen CS, Torsheim T, Brunborg GS, Pallesen S. Development of a Facebook addiction scale. Psychol Rep 2012 Apr;110(2):501-517. [doi: 10.2466/02.09.18.PR0.110.2.501-517] [Medline: 22662404]
25. Wang H, Sigerson L, Cheng C. Digital nativity and information technology addiction: age cohort versus individual difference approaches. Comput Human Behav 2019 Jan;90:1-9. [doi: 10.1016/j.chb.2018.08.031]
26. Kendell R. Diagnostic and statistical manual of mental disorders, 3rd edition. Am J Psychiatry 1988 Oct;145(10):1301-1302 [FREE Full text] [doi: 10.1176/ajp.145.10.1301]
27. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). Washington, DC: American Psychiatric Publishing; 2013.
28. Brown R. Gaming, gambling and other addictive play. In: Adult Play. Amsterdam, the Netherlands: Swets and Zillinger; 1991.
29. Pontes HM, Taylor M, Stavropoulos V. Beyond "Facebook Addiction": the role of cognitive-related factors and psychiatric distress in social networking site addiction. Cyberpsychol Behav Soc Netw 2018 Apr;21(4):240-247. [doi: 10.1089/cyber.2017.0609] [Medline: 29589972]
30. Alaika O, Doghmi N, Cherti M. Social media addiction among Moroccan university students: a cross sectional survey. PAMJ-OH 2020 Apr 3;1:4. [doi: 10.11604/pamj-oh.2020.1.4.21930]
31. Mamun MA, Griffiths MD. The association between Facebook addiction and depression: a pilot survey study among Bangladeshi students. Psychiatry Res 2019 Jan;271:628-633. [doi: 10.1016/j.psychres.2018.12.039] [Medline: 30791335]
32. Azizi SM, Soroush A, Khatony A. The relationship between social networking addiction and academic performance in Iranian students of medical sciences: a cross-sectional study. BMC Psychol 2019 May 03;7(1):28 [FREE Full text] [doi: 10.1186/s40359-019-0305-0] [Medline: 31053171]
33. Brailovskaia J, Schillack H, Margraf J. Facebook addiction disorder in Germany. Cyberpsychol Behav Soc Netw 2018 Jul;21(7):450-456. [doi: 10.1089/cyber.2018.0140] [Medline: 29995531]
34. Tang CS, Koh YY. Online social networking addiction among college students in Singapore: comorbidity with behavioral addiction and affective disorder. Asian J Psychiatr 2017 Feb;25:175-178. [doi: 10.1016/j.ajp.2016.10.027] [Medline: 28262144]
35. Bujang MA, Adnan TH. Requirements for minimum sample size for sensitivity and specificity analysis. J Clin Diagn Res 2016 Oct;10(10):YE01-YE06 [FREE Full text] [doi: 10.7860/JCDR/2016/18129.8744] [Medline: 27891446]
36. Oberski D. Mixture models: latent profile and latent class analysis. In: Modern Statistical Methods for HCI. Cham: Springer; 2016.
37. Collins L, Lanza S. Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences. Hoboken, NJ: John Wiley and Sons; 2009.
38. Williams G, Kibowski F. Latent class analysis and latent profile analysis. In: Handbook of Methodological Approaches to Community-Based Research: Qualitative, Quantitative, and Mixed Methods. New York, NY: Oxford University Press; 2016.

39. Bányai F, Zsila A, Király O, Maraz A, Elekes Z, Griffiths MD, et al. Problematic social media use: results from a large-scale nationally representative adolescent sample. PLoS One 2017 Jan 9;12(1):e0169839 [FREE Full text] [doi: 10.1371/journal.pone.0169839] [Medline: 28068404]
40. Robinson A, Bonnette A, Howard K, Ceballos N, Dailey S, Lu Y, et al. Social comparisons, social media addiction, and social interaction: an examination of specific social media behaviors related to major depressive disorder in a millennial population. J Appl Behav Res 2019 Jan 08;24(1):e12158. [doi: 10.1111/jabr.12158]
41. Keles B, McCrae N, Grealish A. A systematic review: the influence of social media on depression, anxiety and psychological distress in adolescents. Int J Adol Youth 2019 Mar 21;25(1):79-93. [doi: 10.1080/02673843.2019.1590851]
42. Primack BA, Karim SA, Shensa A, Bowman N, Knight J, Sidani JE. Positive and negative experiences on social media and perceived social isolation. Am J Health Promot 2019 Jul;33(6):859-868. [doi: 10.1177/0890117118824196] [Medline: 30661371]
43. Feinstein BA, Bhatia V, Hershenberg R, Davila J. Another venue for problematic interpersonal behavior: the effects of depressive and anxious symptoms on social networking experiences. J Soc Clin Psychol 2012 Apr;31(4):356-382. [doi: 10.1521/jscp.2012.31.4.356]
44. Rosenthal SR, Buka SL, Marshall BD, Carey KB, Clark MA. Negative experiences on Facebook and depressive symptoms among young adults. J Adolesc Health 2016 Nov;59(5):510-516. [doi: 10.1016/j.jadohealth.2016.06.023] [Medline: 27546886]
45. Sigerson L, Li AY, Cheung MW, Cheng C. Examining common information technology addictions and their relationships with non-technology-related addictions. Comput Human Behav 2017 Oct;75:520-526. [doi: 10.1016/j.chb.2017.05.041]
46. Ruggieri S, Santoro G, Pace U, Passanisi A, Schimmenti A. Problematic Facebook use and anxiety concerning use of social media in mothers and their offspring: an actor-partner interdependence model. Addict Behav Rep 2020 Jun;11:100256 [FREE Full text] [doi: 10.1016/j.abrep.2020.100256] [Medline: 32467845]
47. Lee-Won RJ, Herzog L, Park SG. Hooked on Facebook: the role of social anxiety and need for social assurance in problematic use of Facebook. Cyberpsychol Behav Soc Netw 2015 Oct;18(10):567-574. [doi: 10.1089/cyber.2015.0002] [Medline: 26383178]
48. Casale S, Fioravanti G. Satisfying needs through Social Networking Sites: a pathway towards problematic internet use for socially anxious people? Addict Behav Rep 2015 Jun;1:34-39 [FREE Full text] [doi: 10.1016/j.abrep.2015.03.008] [Medline: 29531977]
49. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. J Experimental Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
50. World internet usage and population statistics 2020 year-Q3 estimates. Internet World Stats. URL: https://www.internetworldstats.com/stats.htm [accessed 2020-12-09]
51. Cheng C, Jose PE, Sheldon KM, Singelis TM, Cheung MW, Tiliouine H, et al. Sociocultural differences in self-construal and subjective well-being: a test of four cultural models. J Cross Cult Psychol 2011 Jul 1;42(5):832-855. [doi: 10.1177/0022022110381117]
52. Wang H, Cheng C. Psychometric evaluation and comparison of two gaming disorder measures derived from the DSM-5 and ICD-11 frameworks. Front Psychiatry 2020 Dec 17;11:577366 [FREE Full text] [doi: 10.3389/fpsyt.2020.577366] [Medline: 33391047]
53. Shensa A, Escobar-Viera CG, Sidani JE, Bowman ND, Marshal MP, Primack BA. Problematic social media use and depressive symptoms among U.S. young adults: a nationally-representative study. Soc Sci Med 2017 Jun;182:150-157 [FREE Full text] [doi: 10.1016/j.socscimed.2017.03.061] [Medline: 28446367]
54. Pontes HM. Investigating the differential effects of social networking site addiction and internet gaming disorder on psychological health. J Behav Addict 2017 Dec 01;6(4):601-610 [FREE Full text] [doi: 10.1556/2006.6.2017.075] [Medline: 29130329]
55. Radloff LS. The CES-D Scale: a self-report depression scale for research in the general population. Applied Psychol Measurement 1977 Jun 01;1(3):385-401. [doi: 10.1177/014662167700100306]
56. Spielberger C. Manual for the State-Trait Anxiety Inventory (Form Y). Palo Alto, CA: Consulting Psychologists; 1983.
57. Ekkekakis P, Zenko Z. Measurement of affective responses to exercise: from “affectless arousal” to “the most well-characterized” relationship between the body and affect. In: Emotion Measurement. Cambridge, MA: Woodhead; 2016.
58. Ferguson SL, G. Moore EW, Hull DM. Finding latent groups in observed data: a primer on latent profile analysis in Mplus for applied researchers. Int J Behav Develop 2019 Nov 22;44(5):458-468. [doi: 10.1177/0165025419881721]
59. Tein J, Coxe S, Cham H. Statistical power to detect the correct number of classes in latent profile analysis. Struct Equ Modeling 2013 Oct 01;20(4):640-657 [FREE Full text] [doi: 10.1080/10705511.2013.824781] [Medline: 24489457]
60. Trevethan R. Sensitivity, specificity, and predictive values: foundations, pliabilities, and pitfalls in research and practice. Front Public Health 2017 Nov 20;5:307 [FREE Full text] [doi: 10.3389/fpubh.2017.00307] [Medline: 29209603]
61. Cheng C, Lau Y, Chan L, Luk JW. Prevalence of social media addiction across 32 nations: meta-analysis with subgroup analysis of classification schemes and cultural values. Addict Behav 2021 Jun;117:106845 [FREE Full text] [doi: 10.1016/j.addbeh.2021.106845] [Medline: 33550200]

62. Ettman CK, Abdalla SM, Cohen GH, Sampson L, Vivier PM, Galea S. Prevalence of depression symptoms in US adults before and during the COVID-19 pandemic. JAMA Netw Open 2020 Sep 01;3(9):e2019686 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.19686] [Medline: 32876685]
63. O'Connor RC, Wetherall K, Cleare S, McClelland H, Melson AJ, Niedzwiedz CL, et al. Mental health and well-being during the COVID-19 pandemic: longitudinal analyses of adults in the UK COVID-19 Mental Health and Wellbeing study. Br J Psychiatry 2020 Oct 21;218:326-333 [FREE Full text] [doi: 10.1192/bjp.2020.212] [Medline: 33081860]
64. Cho C. South Korea's efforts to prevent Internet addiction. In: Cambridge Handbook of International Prevention Science. Cambridge, UK: Cambridge University; 2016.
65. Landau S, Milich R, Widiger TA. Predictive power methods may be more helpful for making a diagnosis than sensitivity and specificity. J Child Adolesc Psychopharmacol 1991;1(5):343-351. [doi: 10.1089/cap.1991.1.343] [Medline: 19630684]
66. Streiner DL. Diagnosing tests: using and misusing diagnostic and screening tests. J Pers Assess 2003 Dec;81(3):209-219. [doi: 10.1207/S15327752JPA8103_03] [Medline: 14638445]
67. Petscher Y, Kim Y, Foorman BR. The importance of predictive power in early screening assessments: implications for placement in the response to intervention framework. Assess Eff Interv 2011 Jun;36(3):158-166 [FREE Full text] [doi: 10.1177/1534508410396698] [Medline: 26346970]
68. Ghali H, Ghammem R, Zammit N, Fredj S, Ammari F, Maatoug J, et al. Validation of the Arabic version of the Bergen Facebook addiction scale in Tunisian adolescents. Int J Adolesc Med Health 2019 Sep 24 (forthcoming). [doi: 10.1515/ijamh-2019-0077] [Medline: 31550234]
69. Cantor N, Genero N. Psychiatric diagnosisnatural categorization: A close analogy. In: Contemporary Directions in Psychopathology: Toward the DSM-IV. New York, NY: Guilford; 1986.
70. International Statistical Classification of Diseases and Related Health Problems. 11th edition. World Health Organization. URL: https://www.who.int/standards/classifications/classification-of-diseases [accessed 2021-12-13]
71. Petry NM, Rehbein F, Gentile DA, Lemmens JS, Rumpf H, Mößle T, et al. An international consensus for assessing internet gaming disorder using the new DSM-5 approach. Addiction 2014 Sep;109(9):1399-1406. [doi: 10.1111/add.12457] [Medline: 24456155]
72. Brand M, Rumpf H, Demetrovics Z, Müller A, Stark R, King D, et al. Which conditions should be considered as disorders in the International Classification of Diseases (ICD-11) designation of "other specified disorders due to addictive behaviors"? J Behav Addict 2020 Jun 30 (forthcoming). [doi: 10.1556/2006.2020.00035] [Medline: 32634114]
73. Choi H, Marks NF. Marital conflict, depressive symptoms, and functional impairment. J Marriage Fam 2008;70(2):377-390 [FREE Full text] [doi: 10.1111/j.1741-3737.2008.00488.x] [Medline: 18698378]
74. Sheeber L, Hops H, Alpert A, Davis B, Andrews J. Family support and conflict: prospective relations to adolescent depression. J Abnorm Child Psychol 1997 Aug;25(4):333-344. [doi: 10.1023/a:1025768504415] [Medline: 9304449]
75. Cheng C, Hui W, Lam S. Psychosocial factors and perceived severity of functional dyspeptic symptoms: a psychosocial interactionist model. Psychosom Med 2004;66(1):85-91. [doi: 10.1097/01.psy.0000106885.40753.c1] [Medline: 14747642]
76. Ryum T, Kennair LE, Hjemdal O, Hagen R, Halvorsen J, Solem S. Worry and metacognitions as predictors of anxiety symptoms: a prospective study. Front Psychol 2017 May 31;8:924 [FREE Full text] [doi: 10.3389/fpsyg.2017.00924] [Medline: 28620338]
77. Barlow DH. Anxiety and Its Disorders: The Nature and Treatment of Anxiety and Panic. 2nd edition. New York, NY: Guilford; 2004.
78. Ottenbreit ND, Dobson KS. Avoidance and depression: the construction of the Cognitive–Behavioral Avoidance Scale. Behav Res Ther 2004 Mar;42(3):293-313. [doi: 10.1016/s0005-7967(03)00140-2]
79. Maciosek MV, Coffield AB, Edwards NM, Flottemesch TJ, Goodman MJ, Solberg LI. Priorities among effective clinical preventive services: results of a systematic review and analysis. Am J Prev Med 2006 Jul;31(1):52-61. [doi: 10.1016/j.amepre.2006.03.012] [Medline: 16777543]
80. Adhabi EA, Anozie CB. Literature review for the type of interview in qualitative research. Int J Educ 2017 Sep 20;9(3):86-97. [doi: 10.5296/ije.v9i3.11483]
81. Willis R. The use of composite narratives to present interview findings. Qual Res 2018 Jul 20;19(4):471-480. [doi: 10.1177/1468794118787711]
82. Laconi S, Rodgers RF, Chabrol H. The measurement of Internet addiction: a critical review of existing scales and their psychometric properties. Comput Human Behav 2014 Dec;41:190-202. [doi: 10.1016/j.chb.2014.09.026]
83. Sigerson L, Cheng C. Scales for measuring user engagement with social network sites: a systematic review of psychometric properties. Comput Human Behav 2018 Jun;83:87-105. [doi: 10.1016/j.chb.2018.01.023]
84. Cheng C, Cheung MW, Wang H. Multinational comparison of internet gaming disorder and psychosocial problems versus well-being: meta-analysis of 20 countries. Comput Human Behav 2018 Nov;88:153-167. [doi: 10.1016/j.chb.2018.06.033]
85. Ho RC, Zhang MW, Tsang TY, Toh AH, Pan F, Lu Y, et al. The association between internet addiction and psychiatric co-morbidity: a meta-analysis. BMC Psychiatry 2014 Jun 20;14:183 [FREE Full text] [doi: 10.1186/1471-244X-14-183] [Medline: 24947851]

86. Alimoradi Z, Lin C, Imani V, Griffiths MD, Pakpour AH. Social media addiction and sexual dysfunction among Iranian women: the mediating role of intimacy and social support. J Behav Addict 2019 Jun 01;8(2):318-325 [FREE Full text] [doi: 10.1556/2006.8.2019.24] [Medline: 31120317]
87. Benson V, Hand C, Hartshorne R. How compulsive use of social media affects performance: insights from the UK by purpose of use. Behav Inf Technol 2018 Oct 26;38(6):549-563. [doi: 10.1080/0144929x.2018.1539518]
88. Cheng C, Ebrahimi OV, Lau Y. Maladaptive coping with the infodemic and sleep disturbance in the COVID-19 pandemic. J Sleep Res 2021 Aug;30(4):e13235 [FREE Full text] [doi: 10.1111/jsr.13235] [Medline: 33247519]
89. Brailovskaia J, Margraf J. The relationship between burden caused by coronavirus (Covid-19), addictive social media use, sense of control and anxiety. Comput Human Behav 2021 Jun;119:106720 [FREE Full text] [doi: 10.1016/j.chb.2021.106720] [Medline: 33785982]
90. Cheng C, Wang H, Ebrahimi OV. Adjustment to a "New Normal:" coping flexibility and mental health issues during the COVID-19 pandemic. Front Psychiatry 2021 Mar 19;12:626197 [FREE Full text] [doi: 10.3389/fpsyt.2021.626197] [Medline: 33815166]
91. Cheng C, Cheung MW, Montasem A, 44 members of the International Network of Well-Being Studies. Explaining differences in subjective well-being across 33 nations using multilevel models: universal personality, cultural relativity, and national income. J Pers 2016 Feb;84(1):46-58. [doi: 10.1111/jopy.12136] [Medline: 25234240]

Abbreviations

BSMAS: Bergen Social Media Addiction Scale
DSM: Diagnostic and Statistical Manual of Mental Disorders
ICD: International Statistical Classification of Diseases and Related Health Problems
LPA: latent profile analysis

Edited by R Kukafka; submitted 07.01.21; peer-reviewed by M Attridge, E Said-Hung; comments to author 21.03.21; revised version received 12.04.21; accepted 10.11.21; published 10.01.22.

Please cite as:
Cheng C, Ebrahimi OV, Luk JW
Heterogeneity of Prevalence of Social Media Addiction Across Multiple Classification Schemes: Latent Profile Analysis
J Med Internet Res 2022;24(1):e27000
URL: https://www.jmir.org/2022/1/e27000
doi:10.2196/27000
PMID:35006084

Review

Effectiveness of Live Health Professional–Led Group eHealth Interventions for Adult Mental Health: Systematic Review of Randomized Controlled Trials

Cheryl L Currie[1], PhD; Richard Larouche[1], PhD; M Lauren Voss[1], MSc; Maegan Trottier[1], BA; Rae Spiwak[2], PhD; Erin Higa[1], BSc; David R Scott[3], MLiS; Treena Tallow[4], BA

[1]Faculty of Health Sciences, University of Lethbridge, Lethbridge, AB, Canada

[2]Department of Surgery, Max Rady College of Medicine, University of Manitoba, Winnipeg, MB, Canada

[3]Library Services, University of Lethbridge, Lethbridge, AB, Canada

[4]Alberta Health Services, Lethbridge, AB, Canada

Corresponding Author:
Cheryl L Currie, PhD
Faculty of Health Sciences
University of Lethbridge
4401 University Drive
Lethbridge, AB, T1K 3M4
Canada
Phone: 1 4033324060
Email: cheryl.currie@uleth.ca

Abstract

Background: The COVID-19 pandemic has had adverse impacts on mental health and substance use worldwide. Systematic reviews suggest eHealth interventions can be effective at addressing these problems. However, strong positive eHealth outcomes are often tied to the intensity of web-based therapist guidance, which has time and cost implications that can make the population scale-up of more effective interventions difficult. A way to offset cost while maintaining the intensity of therapist guidance is to offer eHealth programs to groups rather than more standard one-on-one formats.

Objective: This systematic review aims to assess experimental evidence for the effectiveness of live health professional–led group eHealth interventions on mental health, substance use, or bereavement among community-dwelling adults. Within the articles selected for our primary aim, we also seek to examine the impact of interventions that encourage physical activity compared with those that do not.

Methods: Overall, 4 databases (MEDLINE, CINAHL, PsycINFO, and the Cochrane Library) were searched in July 2020. Eligible studies were randomized controlled trials (RCTs) of eHealth interventions led by health professionals and delivered entirely to adult groups by videoconference, teleconference, or webchat. Eligible studies reported mental health, substance use, or bereavement as primary outcomes. The results were examined by outcome, eHealth platform, and intervention length. Postintervention data were used to calculate effect size by study. The findings were summarized using the Synthesis Without Meta-Analysis guidelines. Risk of bias was assessed using the Cochrane Collaboration Tool.

Results: Of the 4099 identified studies, 21 (0.51%) RCTs representing 20 interventions met the inclusion criteria. These studies examined mental health outcomes among 2438 participants (sample size range: 47-361 participants per study) across 7 countries. When effect sizes were pooled, live health professional–led group eHealth interventions had a medium effect on reducing anxiety compared with inactive (Cohen *d*=0.57) or active control (Cohen *d*=0.48), a medium to small effect on reducing depression compared with inactive (Cohen *d*=0.61) or active control (Cohen *d*=0.21), and mixed effects on mental distress and coping. Interventions led by videoconference, and those that provided 8-12 hours of live health professional–led group contact had more robust effects on adult mental health. Risk of bias was high in 91% (19/21) of the studies. Heterogeneity across interventions was significant, resulting in low to very low quality of evidence. No eligible RCT was found that examined substance use, bereavement, or physical activity.

Conclusions: Live eHealth group interventions led by health professionals can foster moderate improvements in anxiety and moderate to small improvements in depression among community-based adults, particularly those delivered by videoconference and those providing 8-12 hours of synchronous engagement.

Trial Registration: PROSPERO CRD42020187551; https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=187551
International Registered Report Identifier (IRRID): RR2-10.1186/s13643-020-01479-3

(J Med Internet Res 2022;24(1):e27939) doi:10.2196/27939

KEYWORDS

systematic review; telemedicine; eHealth; mHealth; e-therapy; mobile interventions; internet; adult; mental health; anxiety; depression; substance use; bereavement; physical activity; CBT; psychotherapy; group; synchronous; videoconference; teleconference

Introduction

Background

The COVID-19 pandemic has resulted in many forms of loss. These privations have had adverse impacts on adult mental health and substance use worldwide [1-3]. Scalable interventions are needed to address the immediate-, mid-, and long-term psychological consequences of the pandemic across countries [4]. Mental health interventions delivered by video, telephone, and mobile apps, termed eHealth interventions, are critical and effective tools to address this need [5-12].

When interventions are delivered using eHealth platforms, patient engagement in the intervention and clinical outcomes are often tied to the intensity of the interaction between clinicians and their clients [8]. For example, a systematic review found that one-on-one therapist-supported internet cognitive behavioral therapy (CBT) was more effective than attention, information, or web-based discussion group controls at reducing adult anxiety [13]. Similarly, a 2021 randomized controlled trial (RCT) found that live, therapist-guided web-based therapy was superior to an unguided form of the program delivered by email at improving pandemic-induced anxiety and depression [14]. A review of best practices to improve engagement and adherence in eHealth interventions specifically recommends the inclusion of human, or a sense of human, contact to build a therapeutic alliance with clients and help them feel more accountable to engage in the interventions [15].

A pair of 2019 systematic reviews concluded that health professional–led videoconference interventions are effective at improving adult anxiety and depression when delivered live and one-on-one between a therapist and a client [11,12]. Yet, therapist engagement in eHealth interventions has time and cost implications that can make the population scale-up and accessibility of more effective programs difficult [16]. A way to offset time and cost while maintaining the intensity of therapist guidance is to offer live therapist-led eHealth interventions to *groups* rather than one-on-one. If effective, group-based eHealth interventions could expand public access to mental health professionals during and after the pandemic.

In Canada and other countries, there is also a need for eHealth interventions that are accessible and culturally safe for Indigenous people and remote communities in addition to the general population [17]. Group-delivered interventions may be more in keeping with the community-centered focus of many Indigenous cultures and thus may be more culturally appropriate for Indigenous clients and communities [18,19]. Group-delivered mental health and substance use interventions have been shown to have cultural utility for Indigenous people, while also providing benefits for non-Indigenous people seeking help for these problems [20,21]. This is not surprising, given that humans are social creatures. Our lives are shaped by our experiences in groups, making the propensity to congregate a powerful therapeutic tool [22]. Group interventions have been shown to promote client engagement in treatment through rewarding and therapeutic forces such as affiliation, support, empathy, and identification [22]. Group interventions can also enable those struggling with mental health and addiction to witness and strive for the healing they see in others, as well as reduce the sense of isolation that mental health and addiction problems can create [23-25].

The need for a systematic review to understand whether group health professional–led interventions could be delivered effectively using eHealth platforms became apparent to our research team in 2020. The year before, we had launched an RCT to assess the mental health impacts of health professional–led interventions delivered to groups in person [26]. In March 2020, we stopped the RCT abruptly because of rapidly spreading COVID-19 in our region and public health restrictions on indoor gatherings. Our in-person mental health interventions had been carefully designed over many months using the principles of patient-oriented research, defined as a process that engages patients and providers, focuses on patient-relevant priorities, and seeks to improve health care practices to improve patient outcomes [27]. When pandemic restrictions required that our interventions move to the web, our Indigenous, patient, and clinical partners recommended that group delivery be maintained. Thus, we sought a systematic review in the literature to guide our efforts. We found 2 systematic reviews that examined this evidence specifically for videoconference-delivered interventions [28,29]. Both concluded that they were feasible for, and well accepted by, adults. Both reviews also observed a trend toward mental health improvement. However, the teams were not confident in their observations, given that they did not specifically search for mental health outcomes. As well, most studies included in these reviews were observational, and interventions delivered by teleconference and live chat platforms were excluded. Thus, we conducted a systematic review of RCTs specifically focused on mental health outcomes for live, health professional–led group interventions delivered by videoconference, teleconference, or live chat platforms both to inform our own work and the work of others.

Review Aims

The primary aim of this systematic review is to assess experimental evidence for the effectiveness of live health

professional–led group eHealth interventions on mental health, substance use, or bereavement among community-dwelling adults. Bereavement and loss were included as outcomes for this review in light of the increased morbidity and mortality that many populations have experienced during the COVID-19 pandemic. A 2020 systematic review of 7 RCTs concluded that web-based one-on-one bereavement interventions are promising [30]. However, the role that group eHealth interventions could play in addressing bereavement outcomes is unknown. Within articles selected for our primary aim, we also sought to examine the impact of eHealth interventions that encouraged physical activity compared with those that did not, given that physical activity has been shown to improve adult mental health [31-33].

Methods

Protocol and Registration

This review was conducted according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement [16,34]. The review protocol was registered with PROSPERO (CRD42020187551) and published as a protocol [35]. Ethics approval and participant consent were not required, given that the review was based on data from previously published studies.

Eligibility Criteria

Eligible Study Designs

This review was limited to RCTs. Nonrandomized and observational studies were excluded. The included studies compared eligible interventions with inactive control interventions (placebo, no treatment, usual or standard care, or a waiting list control) or active control interventions that differed from the treatment intervention (eg, a different variant of the same intervention or a different kind of therapy) in keeping with recommendations from the Cochrane Handbook [36]. Active control interventions included those that were unguided, individual, or delivered in person.

Eligible Interventions

To be eligible, interventions had to be made up of ≥3 sessions delivered entirely live to groups on the web using a video or chat platform or by teleconference. Interventions had to be led by a health facilitator with professional training related to the intervention. This was defined as a certificate or degree in medicine, nursing, allied health, counseling, psychology, social work, or alternative health therapies. Interventions were excluded if they were not delivered in a live, synchronous group format and were not delivered entirely on the web or by telephone. Peer-led groups and web-based groups led by individuals without a recognized certificate or degree related to the intervention were also excluded.

Eligible Participants

Studies that examined community-dwelling adults aged ≥18 years with self-reported or physician-diagnosed mental health, substance use, or bereavement concerns were included. Studies that examined patients in palliative care or adults living in institutionalized settings (eg, care homes, hospitals, and prisons) were excluded.

Eligible Outcomes

The primary outcomes were changes in (1) acute or chronic mental health conditions or concerns, (2) acute or chronic substance use conditions or concerns, or (3) bereavement. In our protocol we had proposed to examine physical health and behavioral outcomes beyond substance use [35]. However, upon reflection we recognized that our search strategy (Multimedia Appendix 1) was not designed to systematically search for these outcomes and we removed them from our review aims. To increase the utility of this review, we have summarized findings from eligible studies that reported physical or behavioral health outcomes other than substance use, with an added caution that these summaries are not based on a comprehensive search of the literature for these outcomes.

Information Sources and Search

In all, 4 electronic databases were searched to identify relevant studies published in English or French from January 2005 to June 2020 (MEDLINE, PsycINFO, CINAHL, and the Cochrane Central Register of Controlled Trials). Reference lists of the selected articles were also searched. The search strategy was developed by a health librarian (DRS) and performed in July 2020 using a combination of key words relevant for each database.

Study Selection and Data Extraction

The results were imported and deduplicated in Covidence (Veritas Health Innovation Ltd) [36]. Titles and abstracts were independently screened in duplicate by 4 reviewers (MLV, EH, MT, and Sydney Murdoch). Full-text screening was conducted by the same reviewers for all articles that met the eligibility criteria or had unclear eligibility. Disagreements were resolved through consensus between 2 reviewers. If a decision could not be reached, consensus was achieved by discussing with an investigator (CLC or RL). Where information in the article was unclear, the corresponding author was contacted for clarification before inclusion. If the corresponding author could not be reached, articles with unclear inclusion criteria were excluded.

Data extraction was carried out independently, in duplicate, by 4 reviewers using Covidence (MLV, EH, MT, and Sydney Murdoch). The extracted data included descriptions of the study sample, intervention details, analytic method, and relationships between the interventions and outcomes of interest. Outcomes reported as means, SDs, and effect estimates were also extracted. Where data were insufficient or not available in the published paper or not obtainable by contacting the authors, studies were excluded from the review. Of note, several RCTs examined the same intervention, reported >1 outcome, or reported >1 measure per outcome, all of which are included in this review.

Synthesis of Results and Assessment of Heterogeneity

Given the expected heterogeneity across studies, it was determined that it would not be appropriate to conduct a meta-analysis. Instead, a narrative analysis of the studies following the Synthesis Without Meta-Analysis reporting guidelines was performed [37]. We explored heterogeneity of the intervention effects by comparing the effect sizes of studies grouped by (1) mental health outcome, (2) intervention delivery platform (videoconference, teleconference, and synchronous

chat), and (3) intervention intensity (<8 contact hours, 8-12 contact hours, and >12 contact hours).

In the protocol, we had planned to assess intervention effects by sex and gender [35]. We were unable to do so because of a lack of studies that reported outcomes by these variables. In the protocol, we had also planned to compare the effects of eHealth interventions delivered to groups with those of eHealth interventions delivered to active or inactive control, with results combined across the 2 control conditions. Upon reviewing the studies selected for the review, we found differential intervention effects for active and inactive controls and have presented them separately in the results.

We found a small but significant subset of studies focused on nonprofessional caregiver mental health. Although we did not specifically search for RCTs related to caregiver mental health, these studies did meet our search criteria. Thus, our main findings summarize results for all participants, including those who are caregivers (eg, parents of children with cancer). To increase the utility of our findings, we have also provided a summary of intervention effects specifically for nonprofessional caregivers.

Quality Assessment

We evaluated the quality of evidence for each outcome using the Grading of Recommendations Assessment, Development and Evaluation approach [38]. Outcomes were assessed using the following categories: (1) *high certainty*: we are very confident that the true effect lies close to the effect estimate; (2) *moderate certainty*: we are moderately confident that the true effect is close to the effect estimate, but there is a possibility that it is substantially different; (3) *low certainty*: we have limited confidence in the effect estimate; the true effect may be substantially different from the effect estimate; and (4) v*ery low certainty*: we have very little confidence in the effect estimate; the true effect is likely to be substantially different from the effect estimate.

Two reviewers (MLV and EH) independently rated the quality of evidence. Given that all studies were RCTs, each rating began as *high quality*. We downgraded quality by 1 level for serious concerns and by 2 levels for very serious concerns about risk of bias, inconsistency, indirectness, imprecision, and publication bias. We used the GRADEpro Guideline Development Tool (McMaster University and Evidence Prime) to generate summary of findings tables for intervention outcomes compared with active and inactive controls [39].

Statistical Analysis

Given that no RCTs selected for this review had a sample size <20 and most had sample sizes >50, we calculated effect sizes using *Cohen d* to allow for a comparison of effects (ie, rather than *Hedges g*, which is typically used to address inflation when sample sizes are <20) [40]. Effect sizes were calculated by subtracting the mean posttest score for the treatment group from the mean posttest score for the control group and dividing the result by the pooled SD of the 2 groups [41]. Effect sizes were computed between the groups within 1 month after the intervention period, given that this time point was most consistently reported across the included studies. Effect sizes were categorized as *trivial* (Cohen d<0.20), *small* (Cohen d=0.20-0.49), *medium* (Cohen d=0.50-0.79), and *large* (Cohen d≥0.80) following the guidelines provided by Cohen [41]. Wilcoxon signed-rank tests in R were used to calculate 95% CIs for pooled effect sizes (with the exception of effect sizes corresponding to single studies, for which 95% CIs could not be produced) [42]. As recommended when a meta-analysis of effect estimates is not conducted, visual displays were created to summarize effect sizes by outcome across studies [43]. Specific effect size calculations by study are provided in Multimedia Appendix 2 [44-64].

Risk of Bias

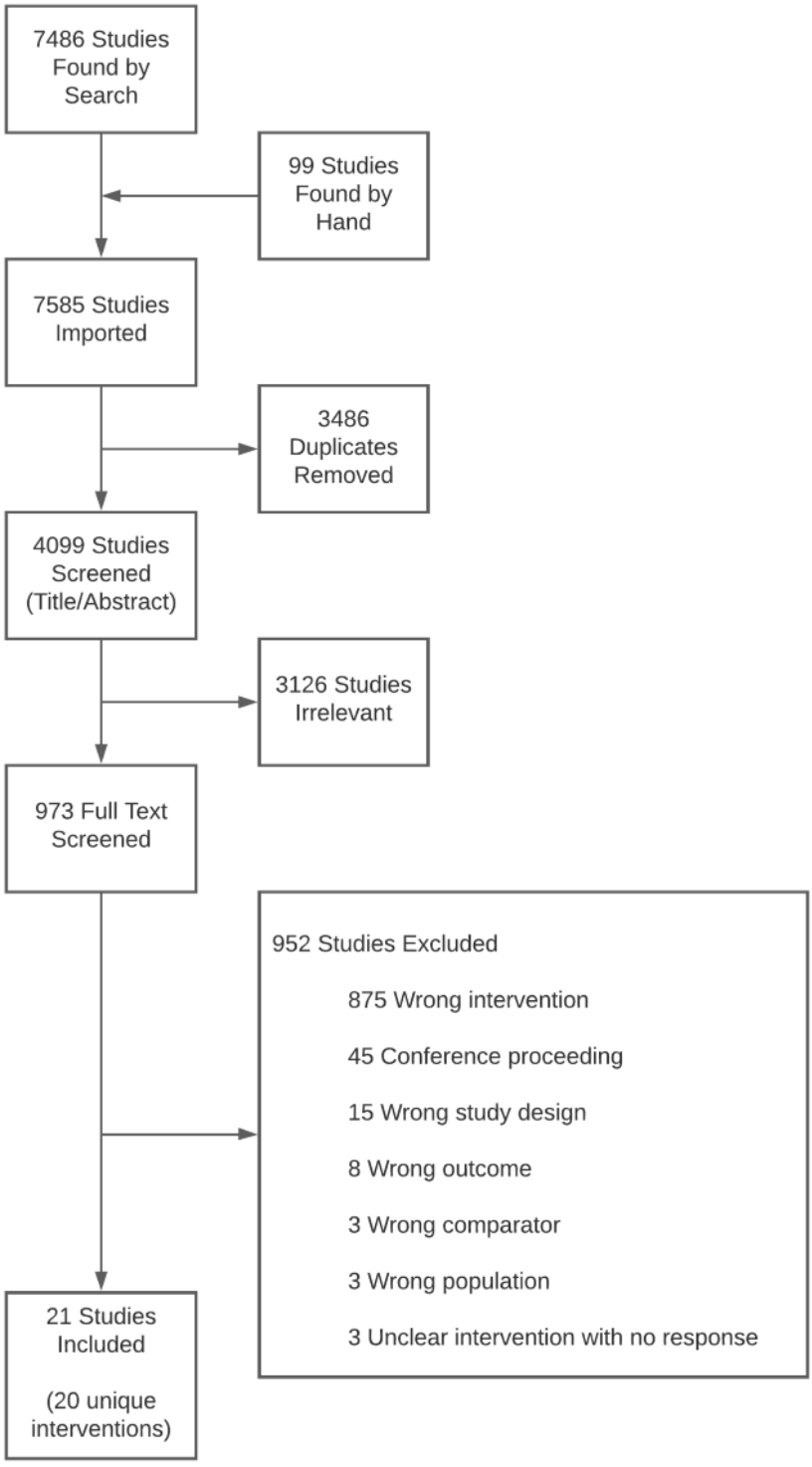
Risk of bias was assessed for all studies using the Cochrane Risk-of-Bias Tool [65]. Of the 21 articles, the first 5 (24%) were assessed by 4 reviewers (MLV, EH, MT, and Sydney Murdoch) and the coinvestigators (CLC, RL, and RS) and compared for consistency. Next, 4 reviewers (MLV, EH, MT, and Sydney Murdoch) independently reviewed articles for risk of bias. Judgments (high, some concerns, and low) were made based on all risk-of-bias domains.

Results

Study Selection

Overall, 21 studies representing 20 interventions were included in this review. As shown in Figure 1, the search yielded 7486 articles, with an additional 99 articles included in the initial screening phase through hand searching the reference lists of identified studies. From the 7585 articles, 3486 (45.96%) duplicates were removed, leaving 4099 (54.04%) articles. After title and abstract screening, of the 4099 articles, 973 (23.74%) remained and underwent a full-text screening by 2 independent reviewers (2 from MLV, EH, MT, and Sydney Murdoch). Within this subsample of 973 articles, 949 (97.5%) were excluded because of ineligibility and 3 (0.3%) were excluded because of unclear eligibility despite multiple efforts to contact authors for clarification.

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



Study and Participant Characteristics

The characteristics of the included studies are summarized in Multimedia Appendix 3 [44-64]. Funding source, outcomes, and measures are summarized in Table 1. A total of 20 interventions delivered to 2438 participants (sample size range: 47 to 361 participants per study) were examined [44-64]. The studies spanned 7 countries, including Canada, the United States, Italy, the United Kingdom, Australia, the Netherlands, and Spain. Most interventions took place in North America.

All 21 studies included in this review assessed changes in mental health conditions or concerns. The most commonly reported outcomes were anxiety, coping, depression, mental distress, and quality of life. None of the included studies reported changes in substance use or bereavement. More than half of the included studies compared interventions using inactive control, which was defined as no intervention, waitlist, or usual care

[44,47,49,52,53,55-58,60,61,63]. This included 2 RCTs (2/21, 10%) that compared 2 interventions that met the inclusion criteria with each other and with inactive control [49,58]. The remaining RCTs in this review compared the intervention with active control, which was defined as any intervention that differed from the treatment and did not meet the eligibility criteria for this review [45,46,51,59,62,64]. Typical active controls were in-person interventions, unguided interventions, and one-on-one interventions. In all, 3 RCTs used both inactive and active control groups [48,50,54].

All studies reported findings by sex, and 14% (3/21) also included gender and sexual orientation (men, women, gay, and bisexual). Although sex was reported in the descriptive summaries, it was not considered during the analyses or the reporting of results, with the exception of Heckman and Carlson [48] who reported findings by male, female, and gay or bisexual categories [48]. Although the studies included in this review were unbalanced in terms of sex, with an overrepresentation of female participants in many RCTs, the findings were typically generalizable to the populations studied. We observed that interventions examining mental health among adults living with HIV had more men, whereas interventions examining mental health among adults with chronic fatigue syndrome had more women. In terms of age, this review focused on adults aged ≥18 years with a single exception. We included a study with participants aged 16-25 years, given that the mean sample age was 20 (SD 2.2) years, with 3.7% (9/244) of the participants aged <18 years [57]. No studies that met our inclusion criteria examined interventions within ethnic minority populations, and no sample was adequate for conducting stratified analysis for ethnic minority adults in this review.

All studies included in this review used self-reported symptom scores to examine the effects of an intervention. No RCT that used physician-diagnosed mental health measures was found. The specific measures used to examine changes in mental health across this review are described in Table 1. Most of the interventions sought to improve mental health among adults with a specific physical health condition, including cancer or tumors [51,59,63,64], HIV or AIDS [47-49], epilepsy [50,55,56], multiple sclerosis (MS) [44,45], bulimia [62], and chronic fatigue [46]. Several sought to improve mental health among nonprofessional caregivers [52,53,58,60,61]. Only 10% (2/20) of interventions that met the inclusion criteria sought to improve mental health among adults in the general population [54,57].

Table 1. Funding source, outcomes, and measures of the included studies (N=21).

Study	Funding source	Outcomes and measures					
		Anxiety	Coping	Depression	Mental distress	Quality of life	Physical health or behavior
Bogosian et al, 2015 [44]	Multiple Sclerosis Society UK (961/11)	HADS[a]	N/A[b]	HADS	GHQ[c]	EQ-5D[d]	MSIS-29[e]; FSS[f]
Cavalera et al, 2019 [45]	Fondazione Italiana Sclerosi Multipla, Italian private foundation (FISM Research Grant 2013/R/17)	HADS	N/A	HADS	N/A	MSQOL-54[g]	MOS-S[h]; MFIS[i]
Hall et al, 2017 [46]	National Institutes of Health (5R01NS055672), National Research Service Award (T32AT000051) from the National Center for Complementary and Integrative Health at the National Institutes of Health	N/A	N/A	N/A	PSS[j]	N/A	CDC-CFS[k]
Heckman et al, 2006 [47]	National Institute on Aging (R21 AG20334)	N/A	CSES[l]; WOCC[m]	GDS[n]	N/A	N/A	N/A
Heckman and Carlson, 2007 [48]	National Institute of Mental Health (RO1 MH59009)	N/A	CSES	BDI[o]	N/A	N/A	N/A
Heckman et al, 2013 [49]	Grant RO1 MH078749 from the National Institute of Mental Health and the National Institute of Nursing Research	N/A	N/A	GDS	N/A	N/A	N/A
Hum et al, 2019 [50]	EpLink: The Epilepsy Research Program of the Ontario Brain Institute	N/A	N/A	QIDS[p] and NDDI-E[q]	N/A	WHOQOL-BREF[r]	N/A
Lepore et al, 2014 [51]	National Institutes of Health (R21CA15877)	HADS	N/A	HADS	N/A	N/A	N/A
Marziali and Donahue, 2006 [52]	National Institute of Mental Health (R34 MH092207)	N/A	N/A	N/A	RMBPC[s]	N/A	HSQ-12[t]
Park et al, 2020 [53]	Marino Health Foundation (no grant number)	PHQ-4[u]	N/A	PHQ-4	VAS[v]	N/A	N/A
Paxton et al, 2007 [54]	Australian Rotary Health Research Fund	N/A	N/A	BDI-II[w]	N/A	N/A	N/A
Thompson et al, 2010 [55]	Cooperative Agreement (U48 DP000043) through the Emory Prevention Research Center from the Centers for Disease Control and Prevention	N/A	CSES	BDI; mBDI[x]; NDDIE; PHQ-9[y]	N/A	BRFSS[z]	N/A
Thompson et al, 2015 [56]	National Institutes of Health grant (5RC1 MD004563) from the National Center for Minority Health and Health Disparities	N/A	N/A	mBDI; PHQ-9	N/A	N/A	N/A
Van der Zanden et al, 2012 [57]	ZonMw (Netherlands Organization for Health Research and Development) grant (61300036)	HADS	N/A	CES-D[aa]	N/A	N/A	N/A
Vazquez et al, 2017 [58]	Ministry of Economy and Competitiveness of Spain (2012-PN162 [PSI2012-37396])	HADS	N/A	CES-D	N/A	N/A	N/A
Vranceanu et al, 2016 [59]; Zale et al, 2018 [64]	Children's Tumor Foundation through a clinical research grant awarded to Ana-Maria Vranceanu	GAS[ab]	MOCS-A[ac]	PHQ-9	N/A	WHOQOL-BREF	NPRS[ad]; BPI[ae]
Wakefield et al, 2016 [60]	Cancer Australia (APP1065428); the National Health and Medical Research Council of Australia (APP1067501); Cancer Institute of New South Wales (11/ECF/3-43); and Cancer Institute of New South Wales (14/ECF/1-11). The Behavioural Sciences Unit is supported by the Kids with Cancer Foundation	DASS-21[af]	N/A	DASS-21	DASS-21	QOL-FCT[ag]	N/A

Study	Funding source	Outcomes and measures					
		Anxiety	Coping	Depression	Mental distress	Quality of life	Physical health or behavior
Winter and Gitlin, 2007 [61]	Alzheimer's Association grant awarded to Laura N. Gitlin, PhD	N/A	N/A	CES-D	N/A	N/A	N/A
Zernicke et al, 2014 [63]	Mind and Life Francisco J. Varela Research Award	POMS[ah]	N/A	POMS	CSOSI[ai]	N/A	N/A
Zerwas et al, 2016 [62]	National Institute of Mental Health grant (R01MH080065); Clinical Translational Science Award (UL1TR000083); and Alexander von Humboldt-Stiftung	BAI[aj]	BDI	N/A	N/A	EDQOL[ak]; SF-6D[al]	N/A

[a]HADS: Hospital Anxiety and Depression Scale.
[b]N/A: not applicable.
[c]GHQ: General Health Questionnaire.
[d]EQ-5D: EuroQol-5 Dimensions.
[e]MSIS-29: Multiple Sclerosis Impact Scale-29.
[f]FSS: Fatigue Severity Scale.
[g]MSQOL-54: Multiple Sclerosis Quality of Life-54.
[h]MOS-S: Medical Outcomes Study-Sleep.
[i]MFIS: Modified Fatigue Impact Scale.
[j]PSS: Perceived Stress Scale.
[k]CDC-CFS: Centers for Disease Control and Prevention Chronic Fatigue Syndrome Symptom Inventory.
[l]CSES: Coping Self-Efficacy Scale.
[m]WOCC: Ways of Coping Checklist.
[n]GDS: Geriatric Depression Scale.
[o]BDI: Beck Depression Inventory.
[p]QIDS: Quick Inventory of Depressive Symptomatology.
[q]NDDI-E: Neurological Disorders Depression Inventory for Epilepsy.
[r]WHOQOL-BREF: World Health Organization Quality of Life-Brief Version.
[s]RMBPC: Revised Memory and Behavior Problems Checklist.
[t]HSQ-12: Health Status Questionnaire-12.
[u]PHQ-4: Patient Health Questionnaire-4 item.
[v]VAS: Visual Analog Scale.
[w]BDI-II: Beck Depression Inventory 2.
[x]mBDI: modified Beck Depression Inventory.
[y]PHQ-9: Patient Health Questionnaire-9 item.
[z]BRFSS: Behavioral Risk Factor Surveillance System.
[aa]CES-D: Center for Epidemiological Studies-Depression.
[ab]GAS: Generalized Anxiety Scale.
[ac]MOCS-A: Measure of Current Status-Part A.
[ad]NPRS: Numeric Pain Rating Scale.
[ae]BPI: Brief Pain Inventory.
[af]DASS-21: Depression Anxiety Stress Scale-21.
[ag]QOL-FCT: Quality of Life-Family Caregiver Tool.
[ah]POMS: Profile of Mood States.
[ai]CSOSI: Calgary Symptoms of Stress Inventory.
[aj]BAI: Beck Anxiety Inventory.
[ak]EDQOL: Eating Disorder Quality of Life.
[al]SF-6D: Short Form-6 Dimensions.

Adults with a specific physical health condition were recruited through registries, community-based organizations, health centers, or clinical referrals. People living with cancer or tumors included women with stage I or II breast cancer in the past 36 months [51], patients with neurofibromatosis diagnosed by a medical professional [59,64], and those diagnosed with cancer who completed primary cancer treatment in the last 3 years [63]. Studies that examined people living with HIV or AIDS used

self-reported diagnosis [47-49]. Adults with epilepsy were required to have been diagnosed with the condition for at least one year [50,55], or at least 3 months should have elapsed after the diagnosis at the time of recruitment [56]. Adults with MS were included if they had a diagnosis of primary or secondary progressive MS [44] or a diagnosis of relapsing–remitting or secondary progressive MS, as determined by a neurologist [45]. Zerwas et al [62] used the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, to assess bulimia nervosa [62]. Hall et al [46] required participants to have a physician-determined chronic fatigue syndrome diagnosis based on the definition formulated by Fukuda et al [66].

Intervention Content and Facilitation

Of the 20 group eHealth interventions, 6 (30%) delivered CBT [46,54,57,58,60,62], 1 (5%) was a mindfulness program tested across 2 RCTs [45,63], 4 (20%) used a combination of mindfulness and CBT [44,50,55,56], 2 (10%) were defined as resilience-based programs [53,59,64], and 6 (30%) were support groups led by a health professional [47-49,51,52,61].

In all, 38% (8/21) of RCTs examined interventions delivered by videoconference [44,45,52,53,59,60,63,64], 43% (9/21) by teleconference [46-50,55,56,58,61], and 19% (4/21) by live, synchronous chat room [51,54,57,62]. Of the 20 interventions, 95% (n=19) were fully delivered by facilitators with professional training related to the intervention (the facilitators were typically mental health or allied health professionals). The sole exception had some peer-led delivery but was included, given that the intervention had significantly more contact hours than most interventions examined, and the first 10 hours had been delivered solely by certified health professionals [52].

Intervention Effectiveness by Outcome

Overview

Key outcomes examined across most of the RCTs were depression (17/21, 81% of the studies), anxiety (8/21, 38%), mental distress (6/21, 24%), coping (4/21, 19%), and quality of life (8/21, 38%). None of the RCTs examined substance use, addiction, or bereavement outcomes, highlighting a gap in knowledge regarding the use of live, health professional–led group eHealth interventions for these outcomes. Effect sizes and CIs were calculated for each study to allow for comparison of effects across RCTs, with results summarized in Table 2 and narratively in the next sections. In addition, Figure 2 provides a visual display of the results as recommended when a meta-analysis of effect estimates is not possible [43].

Table 2. Effect of eHealth interventions by outcome and comparator.

Outcome by comparator	Impact	Number of participants (studies)	Certainty of evidence (GRADE[a])
Anxiety			
Inactive control	Four studies had large to small effects, and 1 study had a trivial effect	446 (5 RCTs[b])	Very low[c,d,e]
Active control	Two studies had large to small effects, and 1 study had a trivial effect; 1 study reported inferior results, but effect sizes could not be calculated	380 (4 RCTs)	Very low[c,d,e,f]
Bereavement			
Inactive control	No studies	0 RCTs	N/A[g]
Active control	No studies	0 RCTs	N/A
Coping			
Inactive control	One study had a small effect; and 1 study showed a trivial effect; 1 study had small to trivial effects favoring the control group	433 (3 RCTs)	Very low[c,d,e]
Active control	One study had a large effect	63 (1 RCTs)	Very low[d,e,f,h]
Depression			
Inactive control	Nine studies had large to small effects, and 2 studies had trivial effects; 1 study comparing 2 interventions found small effects in one and trivial effects in the other. The intervention was inferior to control in 1 study (small effect)	1488 (13 RCTs)	Low[c]
Active control	Three studies had medium to small effects, and 2 studies had trivial effects	500 (5 RCTs)	Very low[c,d,e]
Mental distress			
Inactive control	Four studies had large to medium effects, and 1 study had a trivial effect	268 (5 RCTs)	Very low[c,f,h,i]
Active control	The intervention was inferior to control in 1 study (large effect)	100 (1 RCTs)	Very low[d,e,f]
Quality of life			
Inactive control	One study had a small effect, and 2 studies had a trivial effect. The intervention was inferior to control in 1 study (small effect)	268 (4 RCTs)	Very low[c,d,f,j]
Active control	Two studies had large to small effects, and 2 studies had a trivial effect	421 (4 RCTs)	Very low[c,e,i,j]
Substance use			
Inactive control	No studies	0 RCTs	N/A
Active control	No studies	0 RCTs	N/A

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

[b]RCT: randomized controlled trial.

[c]Most articles were rated *high* using the Cochrane Risk-of-Bias Tool.

[d]Magnitude and direction of effect varied across studies.

[e]Variability in how the outcome is measured and the types of interventions.

[f]The total number of participants across studies was small (400 or fewer), and some studies had small improvements, whereas others had nonsignificant results likely because of a small sample size (borderline imprecision).

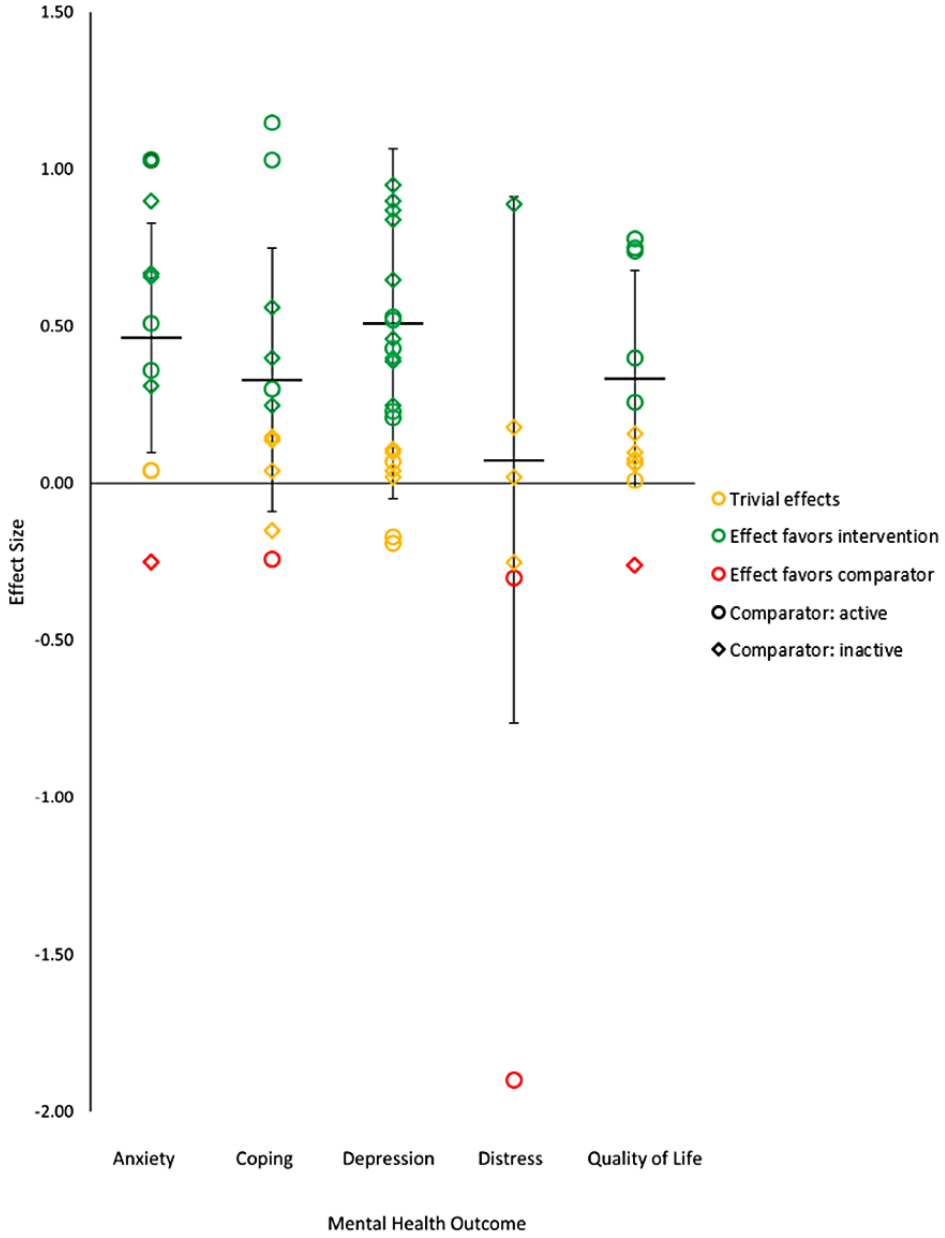
[g]N/A: not applicable.

[h]At least one study was rated *high*, and multiple studies rated *some concerns* overall with the Cochrane Risk-of-Bias Tool.

[i]Populations are limited to a few specific conditions and disorders or by sex, which limits generalizability.

[j]The total number of participants across studies was >400, but some studies found no effect (borderline imprecision).

Figure 2. Effect sizes for mental health outcomes among the included studies compared with inactive and active controls.



Anxiety

In all, 38% (8/21) of RCTs examined intervention effectiveness for anxiety, which was most commonly measured using the Hospital Anxiety and Depression Scale (n=5 studies). The Patient Health Questionnaire (n=1), the Generalized Anxiety Scale (n=1), the Depression Anxiety Stress Scale Short Form (n=1), the Beck Anxiety Inventory (n=1), and the Profile of Mood States (n=1) were also used. Compared with inactive control, health professional–led group eHealth interventions delivered live to adults had large to small effects on anxiety across 4 studies [44,53,57,63] and a trivial effect in 1 study [60]. Compared with active control, the interventions had large to small effects on anxiety in 2 studies [45,59] and a trivial effect in 1 study [62]. Active control was superior to intervention in 1 study; however, we were unable to calculate the effect size using the available data [62]. Taken together, the interventions had a medium effect on reducing anxiety compared with inactive (Cohen *d*=0.57, 95% CI 0.17-0.90) or active control (Cohen *d*=0.48, 95% CI 0.15-0.81).

Coping

A total of 24% (5/21) of studies examined intervention effectiveness for perceived coping. Most of them used the Coping Self-Efficacy Scale (n=3 studies). The Beck Depression Inventory, Measure of Current Status-Part A, and the Ways of Coping Checklist were also used. Compared with inactive control, interventions had a small to trivial effect on coping across 2 studies [47,56]. An additional study compared 2 eligible interventions with inactive control; the effect sizes suggest that the interventions were inferior [48]. Compared with active control, an eligible intervention had a large effect on improving coping (Cohen *d*=1.15, 95% CI values could not be computed) [64]. On the basis of these results, we conclude that more studies are needed before results can be pooled and conclusions drawn about the effectiveness of live health professional–led group eHealth interventions on coping.

Depression

Overall, 81% (17/21) of RCTs examined intervention effectiveness for depression, which was measured using well-recognized instruments for this construct, including the Beck Depression Inventory (n=4 studies), the Hospital Anxiety and Depression Scale (n=3), the Patient Health Questionnaire (n=3), or the Centre of Epidemiological Studies Depression Scale (n=3). Compared with inactive control, 9 interventions had large to small effects on depression [44,50,53,56-58,60,61,63], 1 study had small to trivial effects [49], and 1 study had a trivial effect [47]. Compared with active control, 3 studies had medium to small effects on depression (compared with psychoeducation control) [45,50,59]. However, the intervention was inferior to in-person delivery of the same intervention for depression in 2 studies [54,62]. Taken together, live health professional–led group eHealth interventions had medium effects on adult depression compared with inactive control (Cohen *d*=0.61, 95% CI 0.33-0.89) and small effects on adult depression compared with active control (Cohen *d*=0.21, 95% CI –0.19 to 0.53).

Mental Distress

In all, 24% (5/21) of RCTs examined intervention effectiveness for mental distress, which was examined using a different measure in each of the 5 studies that assessed it. The measures used were the Calgary Symptoms of Stress Inventory, General Health Questionnaire, Perceived Stress Scale, Visual Analog Scale, Revised Memory and Behavior Problems Checklist, and the Depression Anxiety Stress Scale-21 (Table 1). Compared with inactive control, interventions had large to medium effects on mental distress in 4 studies [44,52,53,63] and no effect on mental distress in 1 study [60]. Compared with active control, the intervention was inferior in 1 study compared with in-person delivery (Cohen *d*=–1.90) [46]. When effect sizes were pooled, the eligible interventions yielded a moderate effect on mental distress compared with inactive control (Cohen *d*=0.72, 95% CI –0.04 to 0.85) across 5 studies but inferior to the same intervention delivered in person. Thus, although live health professional–led group eHealth interventions may be effective for addressing mental distress among adults compared with inactive control, more research is needed to determine whether it is effective compared with the same intervention delivered in person.

Quality of Life

This review did not systematically search for quality-of-life outcomes. However, 38% (8/21) of studies that met the inclusion criteria reported intervention effectiveness for this outcome. We have summarized this information, noting that this may not comprehensively summarize all the RCTs that assessed the effect of live eHealth professional–led group interventions on quality of life. As shown in Table 1, quality of life was measured using a variety of instruments. Compared with inactive control, the eligible interventions had a small effect on quality of life in 1 study [44], had a trivial effect in 2 studies [55,56], and were inferior to control in 1 study [60]. Compared with active control, the eligible interventions had large to small effects on quality of life in 2 studies [45,59], and a trivial effect on quality of life in 2 studies [50,62]. When the findings were pooled, we found that live eHealth professional–led group interventions had trivial effects on quality of life compared with inactive control (Cohen *d*=0.07, 95% CI –0.26 to 0.32) and active control (Cohen *d*=0.19, 95% CI 0.07-0.76).

Physical and Behavioral Health

This review did not search for physical or behavioral health outcomes beyond substance use. Given that 24% (5/21) of RCTs selected for this review reported these outcomes, we have summarized the evidence. As shown in Table 1, sleep, fatigue, pain, MS, and general health were examined using a variety of measures. An eligible intervention resulted in significant reductions in fatigue compared with inactive control [44]. However, the results were mixed when compared with active control, with a live health professional–led group eHealth intervention having no effect on fatigue [37] and a second intervention proving inferior to this comparator [38]. For sleep, a study reported significant improvements in sleep quality and sleep quantity compared with active control [37]. For pain, an eligible intervention had a significant impact on pain compared with inactive control [44]. Compared with active control, an intervention reduced pain intensity, but this change did not significantly differ from control at posttest assessment [50]. A study found that a live health professional–led group eHealth intervention significantly reduced the perceived burden of MS on participants' lives compared with inactive control [44]. Finally, a study found that an eligible intervention had no impact on perceived general health compared with no intervention [52].

Effectiveness by Intervention Delivery Platform

Videoconferencing

A total of 38% (8/21) of RCTs examined interventions delivered by videoconference, all of which reported large to small effects on improved mental health compared with inactive [44,52,53,60,63] and active control [45,59,64]. When effect sizes were pooled across comparators, the eligible interventions delivered by videoconferencing had a medium effect on adult anxiety (Cohen *d*=0.60, 95% CI 0.17-1.03), depression (Cohen *d*=0.60, 95% CI 0.38-0.90), and mental distress (Cohen *d*=0.72, 95% CI 0.18-1.03) and a small effect on quality of life (Cohen *d*=0.43, 95% CI –0.03 to 0.99).

Teleconferencing

In all, 43% (9/21) of RCTs examined interventions delivered by teleconference. Compared with inactive control, 6 resulted in large to small improvements in mental health [49,50,55,56,58,61], 1 had a trivial effect [47], and 1 intervention was inferior [48]. Compared with active control, an eligible intervention delivered by teleconference had a small effect on improving mental health in 1 study [40] but was inferior at improving mental health across 2 studies [46,48]. When effect sizes were pooled, teleconference interventions had a medium effect on depression (Cohen *d*=0.50, 95% CI 0.16-1.08), had a trivial effect on coping (Cohen *d*=0.04, 95% CI –0.03 to 0.10) and quality of life (Cohen *d*=0.09, 95% CI 0.06-0.16), and were inferior to control for mental distress (Cohen *d*=–1.90, 95% CI –0.24 to 0.05).

Live Chat Room

Live, synchronous chat rooms were used to deliver interventions across 19% (4/21) of interventions in this review. Compared with inactive control, 2 studies reported large to medium effects on depression and anxiety [54,57]. Compared with active control, 1 study found a trivial effect and 2 studies found that active control was superior (the same intervention delivered face to face) [51,54,62]. When effect sizes were pooled across comparators, interventions delivered by live chat room had a small effect on adult anxiety (Cohen *d*=0.35, 95% CI –0.08 to 0.78) and depression (Cohen *d*=0.24, 95% CI –0.11 to 0.59).

Effectiveness by Intervention Intensity

Contact Hours: <8

Given that only 10% (2/21) of RCTs examined interventions with less than 8 group-based contact hours, we did not pool effect estimates for interventions of this length. Both interventions sought to improve mental health among nonprofessional caregivers. The first intervention had a large effect on depression compared with inactive control and trivial effects compared with active control [58]. The second intervention had small effects on depression, anxiety, and mental distress among parents caring for children with cancer compared with waitlist control [60].

Contact Hours: 8-12

The majority of RCTs (11/21, 52%) examined interventions that had 8-12 group-based contact hours [44,48,50,51,53-57,59,61]. Compared with inactive control, 6 found the interventions superior at improving mental health [44,53-57] and 1 found a trivial effect compared with usual care [48]. Compared with active control, 3 interventions of this length were superior at improving mental health [50,59,64], whereas 2 found that active control was superior [48,54]. When effect sizes were pooled, interventions with 8-12 contact hours had a medium effect on adult anxiety (Cohen *d*=0.57, 95% CI 0.04-1.03) and mental distress (Cohen *d*=0.75, 95% CI 0.05-0.44) and a small effect on coping (Cohen *d*=0.35, 95% CI –0.09 to 0.54) and depression (Cohen *d*=0.40, 95% CI 0.21-0.60).

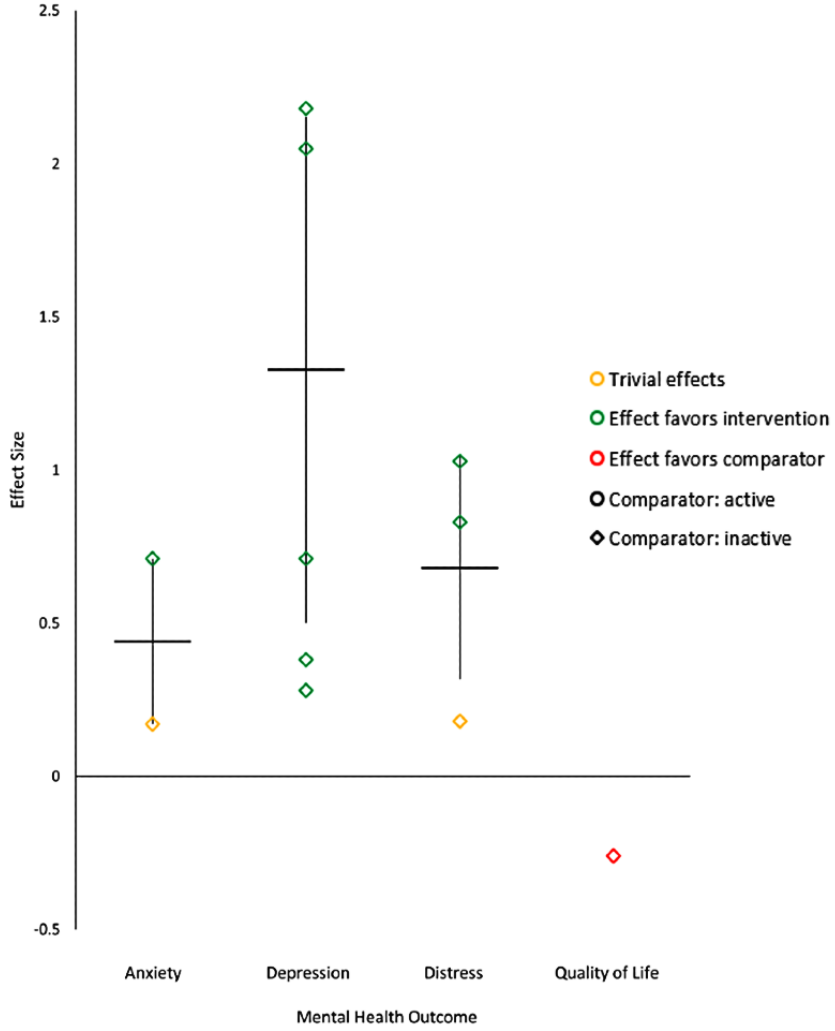
Contact Hours: >12

In all, 38% (8/21) of RCTs examined eligible interventions with more than 12 group-based contact hours [45-47,49,52,61-63]. A total 22 contact hours were provided by 1 study, with the first 10 hours delivered by a health professional and the last 12 hours by peer-led support. We have included this study in the >12-hour category, given that the outcomes were only measured at the end of 22 contact hours [52]. Compared with inactive control, 4 studies found that interventions of this length had large to small effects on mental health outcomes [49,52,61,63] and 1 study found a trivial effect [47]. Compared with active control, 1 study found medium to small effects on mental health [45], 1 study found a trivial effect compared with in-person delivery of the same intervention [62], and 1 study found that in-person delivery was superior [46]. When effect sizes were pooled, interventions with more than 12 contact hours had a medium effect on anxiety (Cohen *d*=0.63, 95% CI 0.26-1.00), a small effect on depression (Cohen *d*=0.33, 95% CI 0.02-0.90), and a trivial effect on coping (Cohen *d*=0.07, 95% CI 0.09-0.21) and mental distress (Cohen *d*=0.01, 95% CI –0.64 to 0.67).

Nonprofessional Caregivers

Almost one quarter of RCTs in this review (5/21, 24%) assessed interventions designed to improve nonprofessional caregiver mental health. These findings were pooled into our overall results. We also present the results for this subgroup separately in Figure 3. To summarize these results, the eligible interventions had beneficial effects on caregiver anxiety [53,60], depression [53,58,60,61], and mental distress [52,53]. Compared with inactive control, 1 study found trivial effects on anxiety [60]. None of the RCTs identified in this review focused on professional caregivers (ie, those trained and paid to provide care), which may be a gap to address in future studies.

Figure 3. Effect sizes for mental health outcomes among caregivers compared with inactive control.



Risk of Bias

Risk-of-bias assessments are presented in Figure 4 and Figure 5 [44-64], with more detailed information presented in Multimedia Appendix 4 [44-64]. Most of the studies (19/21, 91%) received a high risk-of-bias rating, given that the behavioral interventions tested would be difficult to conceal. In all, 19% (4/21) of the RCTs examined interventions that blinded study staff during the data cleaning and analysis portions of the studies [44,51,60,63]. For example, an intervention blinded participants by telling them that they were testing 2 stress management interventions without revealing which was the treatment condition. We included 2 RCTs reporting outcomes from this blinded intervention in this review [59,64]. Risk-of-bias ratings were also affected by the measurement of outcomes, which were typically reported using surveys with participants rather than through independent assessors [67]. Finally, risk-of-bias ratings were affected by missing information about allocation concealment in 67% (14/21) of studies [45-49,52-57,59,61,64] and missing information about the randomization process in 33% (7/21) of studies [46,48,50,52,55,56,61].

The quality of evidence in this review was typically rated very low because of small sample sizes. The exception was depression compared with inactive control, which was rated low given the larger number of studies, larger samples, and medium to large effects (Table 2). For caregiver populations, all outcomes were rated as having very low certainty because of small sample sizes across included studies.

Figure 4. Risk-of-bias graph: review of authors' judgments about each risk-of-bias item presented as percentages across all included studies.

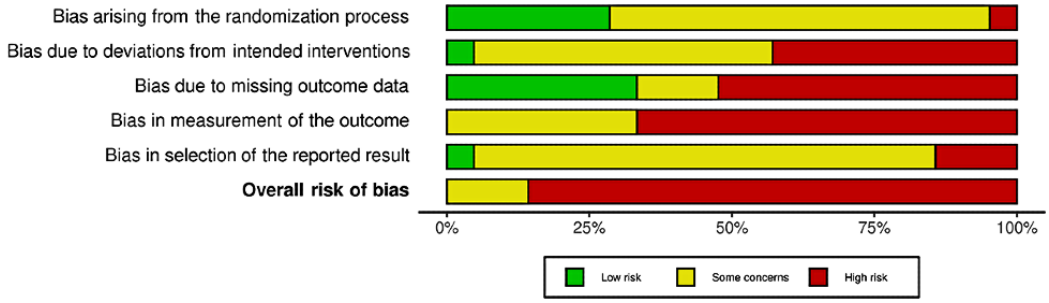
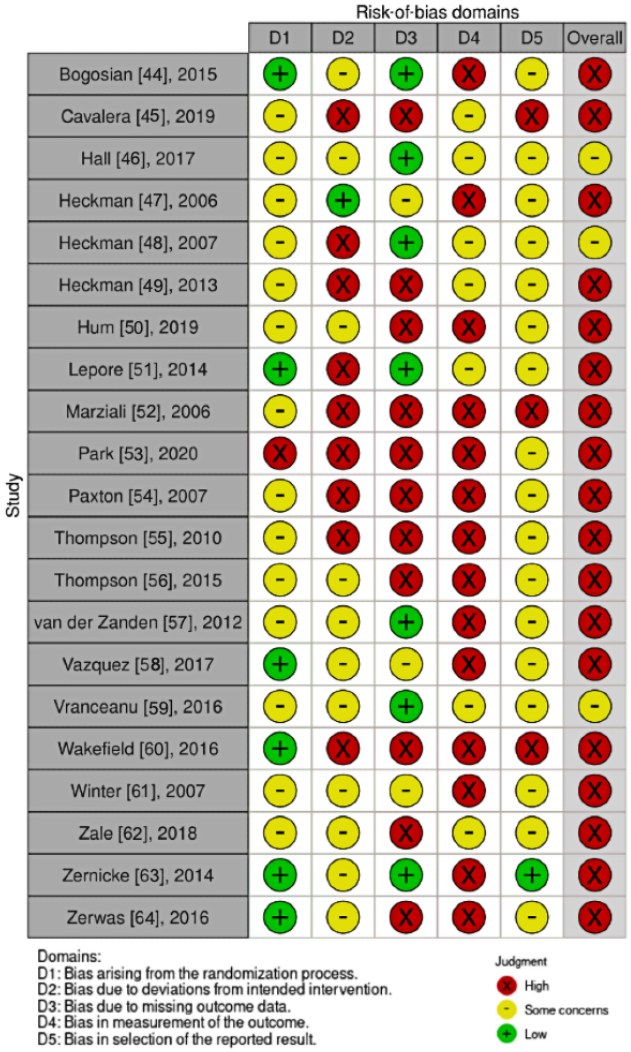


Figure 5. Risk-of-bias summary: review of authors' judgments about each risk-of-bias item for each included study [44-64].



Knowledge Gaps

No eligible RCT that examined impacts on substance use or bereavement was found despite our use of search terms to specifically identify studies that had examined these outcomes. In addition, none of the studies identified in this review reported that they had encouraged physical activity as a way to address or improve mental health as part of their intervention. Finally, no eligible intervention identified in this review focused on professional caregivers or health professionals.

Discussion

Principal Findings

This systematic review assessed experimental evidence for the effectiveness of live health professional–led group eHealth interventions on mental health, substance use, or bereavement among community-dwelling adults. A total of 20 unique interventions met the inclusion criteria for this review across 21 RCTs. Evidence was summarized for 2438 adults in 7 countries. All participants were community-dwelling; most had underlying physical health conditions and were taking part in

the intervention to address mental health concerns related to their health.

The interventions identified in this review had the strongest and most consistent effects on reducing anxiety and depression. Medium effects on anxiety were observed compared with inactive or active control across studies. This is not unexpected, given that eHealth interventions may be particularly effective for anxiety disorders such as agoraphobia that have symptoms exacerbated by travel [11]. For depression, the interventions had medium to small effects across studies compared with inactive and active controls. These findings build on a pair of recent systematic reviews that concluded that live health professional–led videoconference interventions are effective at improving adult anxiety and depression when delivered *one-on-one* [11,12]. Extending these findings to groups in this review is encouraging, given that group delivery of live eHealth interventions could expand public access to mental health professionals in cost- and time-effective ways, both during and after the pandemic. That said, the effectiveness of the interventions examined in this review for mental distress and coping were unclear across studies, highlighting the need for more research on these outcomes.

No RCT was found that examined intervention impacts on substance use or bereavement, highlighting notable gaps in the evidence base. More broadly, systematic reviews of RCTs have shown that internet- and computer-based interventions are effective at reducing grief, as well as cannabis and illicit drug use among adults, although the number of RCTs used to make these assessments remains small [6,30,68]. Thus, it is plausible that live, eHealth group interventions led by health professionals could also be effective at addressing these outcomes. Determining the extent to which this is the case and the superiority of such interventions for these outcomes compared with in-person, asynchronous, or one-on-one eHealth interventions requires further RCT research.

When the results were examined by eHealth platform, interventions delivered by group videoconference had the most robust impacts on adult mental health. Medium effects on reduced anxiety, depression, and mental distress were observed across videoconference interventions, in addition to a small effect on improved quality of life. Eligible teleconference interventions had medium effects on depression. However, trivial effects were observed for improvements in perceived coping and quality of life, and a study found that an intervention delivered by teleconference was inferior to control for mental distress. Although only a few RCTs (n=4) examined interventions delivered using live web-based chat, small but encouraging positive effects were observed for adult anxiety and depression.

In terms of intervention length, 90% (19/21) of the RCTs in this review examined interventions that offered ≥8 live, group contact hours led by a health professional. Interventions with 8-12 contact hours had medium effects on anxiety and mental distress and small effects on depression and coping across 11 RCTs. Interventions with >12 contact hours were assessed by 38% (8/21) of RCTs and also had a medium effect on anxiety and a small effect on depression. This is surprising, given that research suggests that more time is generally better to build a strong therapeutic bond in a digital intervention [15,69]. Building from these results, a hypothesis that could be tested in future studies is whether working with clients in groups using eHealth platforms encourages therapeutic bonds to be built more efficiently compared with more standard one-on-one eHealth therapies. The findings of this review also highlight the need for more research generally, given that the heterogenous nature of the interventions identified, the risk of bias, and the quality of evidence available to date make it difficult to ascertain whether the observed effects were due to intervention length or other factors.

The conflictive results observed in this review may also be due to differences in the target populations examined. Many participants in this review had underlying health conditions, which suggests that some may have been receiving concomitant treatments as part of their usual care that could have influenced outcomes. It may also be that chronic disease or pain among participants in some studies influenced the extent of their engagement. Inconsistent results may also be due to differences in the therapeutic methods used (eg, mindfulness-based cognitive therapy, CBT, and mindfulness-based stress reduction). Of note, we were not able to draw conclusions on the effectiveness of specific therapeutic methods because few studies used the same method. Differences in the types of control groups used across studies may have also influenced the discrepant results. For example, when an active intervention was found to be inferior in this review, it was often in studies that compared it with an active control that was the same intervention delivered face to face. Finally, many of the RCTs included in this review were described as pilot studies (9/21, 43%) and may not have had the study power to draw reliable conclusions.

Limitations

This review was limited to studies published in English or French. The search did not include physical or behavioral health outcomes or gray literature. Clinical experts (eg, psychiatrists) were not involved in the search. All measures used to assess changes in mental health in this review were self-report, and the overall number of studies summarized was small. The nature of the interventions under study (ie, behavioral rather than pharmaceutical) meant that participants were usually unblinded. For these reasons, no study captured in this review was rated lower than *moderate* for risk of bias. Intervention adherence and attendance were not consistently reported across the RCTs selected for this review and could not be assessed. All participants were community-dwelling, and most had underlying chronic health problems. Given that chronic health conditions are common among adults, this may not hamper generalizability, although generalizability to adults without underlying health conditions should be made with caution [57,58]. We also note that the intensity of usual care varied across the RCTs in this review, raising concerns about the interpretation and generalizability of results. We also note that a study identified in this review included 9 participants who were aged 16-17 years. The reason for this exception to the inclusion criteria is that the proportion of the sample who were not aged ≥18 years was extremely small (9/244, 3.8%) and likely inconsequential

to the results, given that the remaining bulk of the sample (235/244, 96.2% of the participants) were aged ≥18 years. Thus, we decided that removing this particular study would be a greater violation of our inclusion criteria than including it would be a violation of our exclusion criteria. This study had one of the largest sample sizes of adults who met all inclusion criteria; thus, it was our opinion that removing it would result in a systematic review that did not reflect the published studies we sought to summarize. We can confirm that no other study in the pool of papers selected from our original literature search that met all other inclusion criteria for this review had a mixed adolescent or adult sample. Finally, few studies selected for this review included a follow-up period of ≥6 months. As a result, the longer-term impacts of live health professional–led group eHealth interventions on adult mental health remains unclear.

Conclusion

Live eHealth group interventions led by health professionals can foster moderate improvements in anxiety, and moderate to small improvements in depression among community-dwelling adults, particularly those delivered by videoconference and those providing 8-12 hours of synchronous engagement. This review highlights a need for experimental research to understand the long-term impacts of these interventions and whether they may be effective for adult substance use and bereavement.

Acknowledgments

This work was supported by funding from Alberta Innovates awarded to CLC (#201300491). The authors would like to thank Sydney Murdoch for her assistance in article screening, data extraction, and risk-of-bias assessments.

Conflicts of Interest

None declared.

Multimedia Appendix 1
MEDLINE search strategy.
[PDF File (Adobe PDF File), 87 KB - jmir_v24i1e27939_app1.pdf]

Multimedia Appendix 2
Effect size calculation summary.
[PDF File (Adobe PDF File), 191 KB - jmir_v24i1e27939_app2.pdf]

Multimedia Appendix 3
Characteristics of included studies (N=21).
[PDF File (Adobe PDF File), 214 KB - jmir_v24i1e27939_app3.pdf]

Multimedia Appendix 4
Risk-of-bias screening results.
[PDF File (Adobe PDF File), 116 KB - jmir_v24i1e27939_app4.pdf]

References

1. Currie CL. Adult PTSD symptoms and substance use during Wave 1 of the COVID-19 pandemic. Addict Behav Rep 2021 Jun;13:100341. [doi: 10.1016/j.abrep.2021.100341] [Medline: 33763517]
2. United Nations. COVID-19 and the need for action on mental health. UN Executive Office of the Secretary-General (EOSG) Policy Briefs and Papers. 2020. URL: https://www.un-ilibrary.org/content/papers/27082245/9 [accessed 2021-12-24]
3. Biddle N, Edwards B, Gray M, Sollis K. Alcohol consumption during the COVID19 period: May 2020. Australian National University. 2020. URL: https://csrm.cass.anu.edu.au/sites/default/files/docs/2020/6/Alcohol_consumption_during_the_COVID-19_period.pdf [accessed 2021-12-24]
4. Amsalem D, Dixon LB, Neria Y. The Coronavirus Disease 2019 (COVID-19) outbreak and mental health: current risks and recommended actions. JAMA Psychiatry 2021 Jan 01;78(1):9-10. [doi: 10.1001/jamapsychiatry.2020.1730] [Medline: 32579160]
5. Tait RJ, Spijkerman R, Riper H. Internet and computer based interventions for cannabis use: a meta-analysis. Drug Alcohol Depend 2013 Dec 1;133(2):295-304. [doi: 10.1016/j.drugalcdep.2013.05.012] [Medline: 23747236]
6. Boumparis N, Loheide-Niesmann L, Blankers M, Ebert DD, Korf D, Schaub MP, et al. Short- and long-term effects of digital prevention and treatment interventions for cannabis use reduction: a systematic review and meta-analysis. Drug Alcohol Depend 2019 Jul 01;200:82-94 [FREE Full text] [doi: 10.1016/j.drugalcdep.2019.03.016] [Medline: 31112834]
7. Rooke S, Thorsteinsson E, Karpin A, Copeland J, Allsop D. Computer-delivered interventions for alcohol and tobacco use: a meta-analysis. Addiction 2010 Aug;105(8):1381-1390. [doi: 10.1111/j.1360-0443.2010.02975.x] [Medline: 20528806]

8. Baumeister H, Reichler L, Munzinger M, Lin J. The impact of guidance on Internet-based mental health interventions — a systematic review. Internet Interv 2014 Oct;1(4):205-215. [doi: 10.1016/j.invent.2014.08.003]
9. Titov N. Internet-delivered psychotherapy for depression in adults. Curr Opin Psychiatry 2011 Jan;24(1):18-23. [doi: 10.1097/YCO.0b013e32833ed18f] [Medline: 20827199]
10. Hester RK, Lenberg KL, Campbell W, Delaney HD. Overcoming addictions, a web-based application, and SMART Recovery, an online and in-person mutual help group for problem drinkers, part 1: three-month outcomes of a randomized controlled trial. J Med Internet Res 2013 Jul 11;15(7):e134 [FREE Full text] [doi: 10.2196/jmir.2565] [Medline: 23846588]
11. Berryhill MB, Halli-Tierney A, Culmer N, Williams N, Betancourt A, King M, et al. Videoconferencing psychological therapy and anxiety: a systematic review. Fam Pract 2019 Jan 25;36(1):53-63. [doi: 10.1093/fampra/cmy072] [Medline: 30188992]
12. Berryhill MB, Culmer N, Williams N, Halli-Tierney A, Betancourt A, Roberts H, et al. Videoconferencing psychotherapy and depression: a systematic review. Telemed J E Health 2019 Jun;25(6):435-446. [doi: 10.1089/tmj.2018.0058] [Medline: 30048211]
13. Olthuis JV, Watt MC, Bailey K, Hayden JA, Stewart SH. Therapist-supported internet cognitive behavioural therapy for anxiety disorders in adults. Cochrane Database Syst Rev 2015 Mar 05(3):CD011565. [doi: 10.1002/14651858.CD011565] [Medline: 25742186]
14. Al-Alawi M, McCall RK, Sultan A, Al Balushi N, Al-Mahrouqi T, Al Ghailani A, et al. Efficacy of a six-week-long therapist-guided online therapy versus self-help internet-based therapy for COVID-19-induced anxiety and depression: open-label, pragmatic, randomized controlled trial. JMIR Ment Health 2021 Feb 12;8(2):e26683 [FREE Full text] [doi: 10.2196/26683] [Medline: 33512323]
15. Karekla M, Kasinopoulos O, Neto DD, Ebert DD, Van Daele T, Nordgreen T, et al. Best practices and recommendations for digital interventions to improve engagement and adherence in chronic illness sufferers. Eur Psychol 2019 Jan;24(1):49-67. [doi: 10.1027/1016-9040/a000349]
16. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Br Med J 2015 Jan 02;350:g7647 [FREE Full text] [doi: 10.1136/bmj.g7647] [Medline: 25555855]
17. Curtis E, Jones R, Tipene-Leach D, Walker C, Loring B, Paine S, et al. Why cultural safety rather than cultural competency is required to achieve health equity: a literature review and recommended definition. Int J Equity Health 2019 Nov 14;18(1):174 [FREE Full text] [doi: 10.1186/s12939-019-1082-3] [Medline: 31727076]
18. Currie C, Wild T, Schopflocher D, Laing L, Veugelers P. Illicit and prescription drug problems among urban Aboriginal adults in Canada: the role of traditional culture in protection and resilience. Soc Sci Med 2013 Jul;88:1-9 [FREE Full text] [doi: 10.1016/j.socscimed.2013.03.032] [Medline: 23702204]
19. Smith G, Currie C, Battle J, Belanger Y. Exploring gambling impacts in two Alberta Cree communities: a participatory action study. In: Belanger Y, editor. First Nations gambling in Canada: Current trends and issues. Winnipeg, MB: University of Manitoba Press; 2011:118-139.
20. Dale E, Lee KS, Conigrave KM, Conigrave JH, Ivers R, Clapham K, et al. A multi-methods yarn about SMART Recovery: first insights from Australian Aboriginal facilitators and group members. Drug Alcohol Rev 2021 Sep;40(6):1013-1027 [FREE Full text] [doi: 10.1111/dar.13264] [Medline: 33686719]
21. Lo K, Waterland J, Todd P, Gupta T, Bearman M, Hassed C, et al. Group interventions to promote mental health in health professional education: a systematic review and meta-analysis of randomised controlled trials. Adv Health Sci Educ Theory Pract 2018 May;23(2):413-447. [doi: 10.1007/s10459-017-9770-5] [Medline: 28299475]
22. Substance abuse treatment: group therapy. In: SAMHSA/CSAT Treatment Improvement Protocols. Rockville, MD: Substance Abuse and Mental Health Services Administration (US); 2005.
23. Flores PJ. Group psychotherapy and neuro-plasticity: an attachment theory perspective. Int J Group Psychother 2010 Oct 25;60(4):546-570. [doi: 10.1521/ijgp.2010.60.4.546] [Medline: 21028976]
24. Kösters M, Burlingame G, Nachtigall C, Strauss B. A meta-analytic review of the effectiveness of inpatient group psychotherapy. Group Dyn Theory Res Pract 2006 Jun;10(2):146-163 [FREE Full text] [doi: 10.1037/1089-2699.10.2.146]
25. Schermer VL. Mirror neurons: their implications for group psychotherapy. Int J Group Psychother 2010 Oct 25;60(4):486-513. [doi: 10.1521/ijgp.2010.60.4.486] [Medline: 21028974]
26. Currie CL, Copeland JL, Voss ML, Swanepoel L, Ambeskovic M, Patel NB, et al. Trauma-informed interventions versus control for cancer-risk behaviours among adults: rationale and design for a randomized trial. BMC Public Health 2019 Oct 29;19(1):1403 [FREE Full text] [doi: 10.1186/s12889-019-7641-0] [Medline: 31664968]
27. Strategy for patient-oriented research. Canadian Institutes of Health Research. 2011. URL: https://cihr-irsc.gc.ca/e/41204.html [accessed 2021-12-24]
28. Gentry MT, Lapid MI, Clark MM, Rummans TA. Evidence for telehealth group-based treatment: a systematic review. J Telemed Telecare 2019 Jul;25(6):327-342. [doi: 10.1177/1357633X18775855] [Medline: 29788807]
29. Banbury A, Nancarrow S, Dart J, Gray L, Parkinson L. Telehealth interventions delivering home-based support group videoconferencing: systematic review. J Med Internet Res 2018 Feb 02;20(2):e25 [FREE Full text] [doi: 10.2196/jmir.8090] [Medline: 29396387]

30. Wagner B, Rosenberg N, Hofmann L, Maass U. Web-based bereavement care: a systematic review and meta-analysis. Front Psychiatry 2020 Jun 24;11:525 [FREE Full text] [doi: 10.3389/fpsyt.2020.00525] [Medline: 32670101]
31. Rebar AL, Stanton R, Geard D, Short C, Duncan MJ, Vandelanotte C. A meta-meta-analysis of the effect of physical activity on depression and anxiety in non-clinical adult populations. Health Psychol Rev 2015;9(3):366-378. [doi: 10.1080/17437199.2015.1022901] [Medline: 25739893]
32. Rhodes RE, Janssen I, Bredin SS, Warburton DE, Bauman A. Physical activity: health impact, prevalence, correlates and interventions. Psychol Health 2017 Aug;32(8):942-975. [doi: 10.1080/08870446.2017.1325486] [Medline: 28554222]
33. Robertson R, Robertson A, Jepson R, Maxwell M. Walking for depression or depressive symptoms: a systematic review and meta-analysis. Ment Health Phys Act 2012 Jun;5(1):66-75. [doi: 10.1016/j.mhpa.2012.03.002]
34. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
35. Currie CL, Larouche R, Voss ML, Higa EK, Spiwak R, Scott D, et al. The impact of eHealth group interventions on the mental, behavioral, and physical health of adults: a systematic review protocol. Syst Rev 2020 Sep 23;9(1):217 [FREE Full text] [doi: 10.1186/s13643-020-01479-3] [Medline: 32967717]
36. Covidence systematic review software. Veritas Health Innovation, Melbourne, Australia. URL: https://www.covidence.org/ [accessed 2021-12-24]
37. Campbell M, McKenzie JE, Sowden A, Katikireddi SV, Brennan SE, Ellis S, et al. Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline. Br Med J 2020 Jan 16;368:l6890 [FREE Full text] [doi: 10.1136/bmj.l6890] [Medline: 31948937]
38. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. Br Med J 2008 Apr 26;336(7650):924-926 [FREE Full text] [doi: 10.1136/bmj.39489.470347.AD] [Medline: 18436948]
39. GRADEpro GDT: GRADEpro guideline development tool. McMaster University. 2015. URL: https://www.gradepro.org/ [accessed 2021-12-24]
40. Hedges L, Olkin I. Statistical Methods for Meta-Analysis. Cambridge, MA: Academic Press Inc; 1985.
41. Cohen J. Statistical Power Analysis for the Behavioral Sciences, 2nd Ed. Mahwah, NJ, USA: Lawrence Erlbaum Associates; 1988.
42. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2020. URL: https://www.R-project.org [accessed 2021-12-24]
43. McKenzie J, Brennan S. Synthesizing and presenting findings using other methods. In: Cochrane Handbook for Systematic Reviews of Interventions. Hoboken, New Jersey, United States: John Wiley & Sons; 2019:321-347.
44. Bogosian A, Chadwick P, Windgassen S, Norton S, McCrone P, Mosweu I, et al. Distress improves after mindfulness training for progressive MS: a pilot randomised trial. Mult Scler 2015 Aug;21(9):1184-1194. [doi: 10.1177/1352458515576261] [Medline: 25767124]
45. Cavalera C, Rovaris M, Mendozzi L, Pugnetti L, Garegnani M, Castelnuovo G, et al. Online meditation training for people with multiple sclerosis: a randomized controlled trial. Mult Scler 2019 Apr;25(4):610-617. [doi: 10.1177/1352458518761187] [Medline: 29485319]
46. Hall DL, Lattie EG, Milrad SF, Czaja S, Fletcher MA, Klimas N, et al. Telephone-administered versus live group cognitive behavioral stress management for adults with CFS. J Psychosom Res 2017 Feb;93:41-47 [FREE Full text] [doi: 10.1016/j.jpsychores.2016.12.004] [Medline: 28107891]
47. Heckman TG, Barcikowski R, Ogles B, Suhr J, Carlson B, Holroyd K, et al. A telephone-delivered coping improvement group intervention for middle-aged and older adults living with HIV/AIDS. Ann Behav Med 2006 Aug;32(1):27-38. [doi: 10.1207/s15324796abm3201_4] [Medline: 16827627]
48. Heckman TG, Carlson B. A randomized clinical trial of two telephone-delivered, mental health interventions for HIV-infected persons in rural areas of the United States. AIDS Behav 2007 Jan;11(1):5-14. [doi: 10.1007/s10461-006-9111-9] [Medline: 16705479]
49. Heckman TG, Heckman BD, Anderson T, Lovejoy TI, Mohr D, Sutton M, et al. Supportive-expressive and coping group teletherapies for HIV-infected older adults: a randomized clinical trial. AIDS Behav 2013 Nov;17(9):3034-3044 [FREE Full text] [doi: 10.1007/s10461-013-0441-0] [Medline: 23474642]
50. Hum KM, Chan CJ, Gane J, Conway L, McAndrews MP, Smith ML. Do distance-delivery group interventions improve depression in people with epilepsy? Epilepsy Behav 2019 Sep;98(Pt A):153-160. [doi: 10.1016/j.yebeh.2019.06.037] [Medline: 31374471]
51. Lepore SJ, Buzaglo JS, Lieberman MA, Golant M, Greener JR, Davey A. Comparing standard versus prosocial internet support groups for patients with breast cancer: a randomized controlled trial of the helper therapy principle. J Clin Oncol 2014 Dec 20;32(36):4081-4086. [doi: 10.1200/JCO.2014.57.0093] [Medline: 25403218]
52. Marziali E, Donahue P. Caring for others: internet video-conferencing group intervention for family caregivers of older adults with neurodegenerative disease. Gerontologist 2006 Jun;46(3):398-403. [doi: 10.1093/geront/46.3.398] [Medline: 16731880]

53. Park ER, Perez GK, Millstein RA, Luberto CM, Traeger L, Proszynski J, et al. A virtual resiliency intervention promoting resiliency for parents of children with learning and attentional disabilities: a randomized pilot trial. Matern Child Health J 2020 Jan;24(1):39-53. [doi: 10.1007/s10995-019-02815-3] [Medline: 31650412]
54. Paxton SJ, McLean SA, Gollings EK, Faulkner C, Wertheim EH. Comparison of face-to-face and internet interventions for body image and eating problems in adult women: an RCT. Int J Eat Disord 2007 Dec;40(8):692-704. [doi: 10.1002/eat.20446] [Medline: 17702020]
55. Thompson NJ, Walker ER, Obolensky N, Winning A, Barmon C, Diiorio C, et al. Distance delivery of mindfulness-based cognitive therapy for depression: project UPLIFT. Epilepsy Behav 2010 Nov;19(3):247-254. [doi: 10.1016/j.yebeh.2010.07.031] [Medline: 20851055]
56. Thompson NJ, Patel AH, Selwa LM, Stoll SC, Begley CE, Johnson EK, et al. Expanding the efficacy of Project UPLIFT: distance delivery of mindfulness-based depression prevention to people with epilepsy. J Consult Clin Psychol 2015 Apr;83(2):304-313 [FREE Full text] [doi: 10.1037/a0038404] [Medline: 25495361]
57. van der Zanden R, Kramer J, Gerrits R, Cuijpers P. Effectiveness of an online group course for depression in adolescents and young adults: a randomized trial. J Med Internet Res 2012;14(3):e86 [FREE Full text] [doi: 10.2196/jmir.2033] [Medline: 22677437]
58. Vázquez FL, Torres A, Otero P, Blanco V, Díaz O, Estévez LE. Analysis of the components of a cognitive-behavioral intervention administered via conference call for preventing depression among non-professional caregivers: a pilot study. Aging Ment Health 2017 Sep;21(9):938-946. [doi: 10.1080/13607863.2016.1181714] [Medline: 27187725]
59. Vranceanu A, Riklin E, Merker VL, Macklin EA, Park ER, Plotkin SR. Mind-body therapy via videoconferencing in patients with neurofibromatosis: an RCT. Neurology 2016 Aug 23;87(8):806-814. [doi: 10.1212/WNL.0000000000003005] [Medline: 27449066]
60. Wakefield CE, Sansom-Daly UM, McGill BC, Ellis SJ, Doolan EL, Robertson EG, et al. Acceptability and feasibility of an e-mental health intervention for parents of childhood cancer survivors: "Cascade". Support Care Cancer 2016 Dec;24(6):2685-2694. [doi: 10.1007/s00520-016-3077-6] [Medline: 26781620]
61. Winter L, Gitlin LN. Evaluation of a telephone-based support group intervention for female caregivers of community-dwelling individuals with dementia. Am J Alzheimers Dis Other Demen 2006;21(6):391-397 [FREE Full text] [doi: 10.1177/1533317506291371] [Medline: 17267370]
62. Zerwas SC, Watson HJ, Hofmeier SM, Levine MD, Hamer RM, Crosby RD, et al. CBT4BN: a randomized controlled trial of online chat and face-to-face group therapy for bulimia nervosa. Psychother Psychosom 2017;86(1):47-53 [FREE Full text] [doi: 10.1159/000449025] [Medline: 27883997]
63. Zernicke KA, Campbell TS, Speca M, McCabe-Ruff K, Flowers S, Carlson LE. A randomized wait-list controlled trial of feasibility and efficacy of an online mindfulness-based cancer recovery program: the eTherapy for cancer applying mindfulness trial. Psychosom Med 2014 May;76(4):257-267. [doi: 10.1097/PSY.0000000000000053] [Medline: 24804884]
64. Zale EL, Pierre-Louis C, Macklin EA, Riklin E, Vranceanu A. The impact of a mind-body program on multiple dimensions of resiliency among geographically diverse patients with neurofibromatosis. J Neurooncol 2018 Apr;137(2):321-329. [doi: 10.1007/s11060-017-2720-5] [Medline: 29275505]
65. Sterne JA, Savović J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Br Med J 2019 Aug 28;366:l4898. [doi: 10.1136/bmj.l4898] [Medline: 31462531]
66. Fukuda K, Straus SE, Hickie I, Sharpe MC, Dobbins JG, Komaroff A. The chronic fatigue syndrome: a comprehensive approach to its definition and study. International Chronic Fatigue Syndrome Study Group. Ann Intern Med 1994 Dec 15;121(12):953-959. [doi: 10.7326/0003-4819-121-12-199412150-00009] [Medline: 7978722]
67. Higgins JP, Altman DG, Gøtzsche PC, Jüni P, Moher D, Oxman AD, Cochrane Bias Methods Group, Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. Br Med J 2011 Oct 18;343:d5928 [FREE Full text] [doi: 10.1136/bmj.d5928] [Medline: 22008217]
68. Boumparis N, Karyotaki E, Schaub MP, Cuijpers P, Riper H. Internet interventions for adult illicit substance users: a meta-analysis. Addiction 2017 Sep;112(9):1521-1532 [FREE Full text] [doi: 10.1111/add.13819] [Medline: 28295758]
69. Jasper K, Weise C, Conrad I, Andersson G, Hiller W, Kleinstäuber M. The working alliance in a randomized controlled trial comparing internet-based self-help and face-to-face cognitive behavior therapy for chronic tinnitus. Internet Interv 2014 Apr;1(2):49-57. [doi: 10.1016/j.invent.2014.04.002]

Abbreviations

CBT: cognitive behavioral therapy
MS: multiple sclerosis
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT: randomized controlled trial

Edited by A Mavragani; submitted 21.05.21; peer-reviewed by M Wolf, K Matthias; comments to author 05.08.21; revised version received 03.11.21; accepted 03.12.21; published 11.01.22.

Please cite as:
Currie CL, Larouche R, Voss ML, Trottier M, Spiwak R, Higa E, Scott DR, Tallow T
Effectiveness of Live Health Professional–Led Group eHealth Interventions for Adult Mental Health: Systematic Review of Randomized Controlled Trials
J Med Internet Res 2022;24(1):e27939
URL: https://www.jmir.org/2022/1/e27939
doi:10.2196/27939
PMID:34878409

Original Paper

Examining Tweet Content and Engagement of Users With Tweets About Hikikomori in Japanese: Mixed Methods Study of Social Withdrawal

Victor Pereira-Sanchez[1,2,3,4], MD, PhD; Miguel Angel Alvarez-Mon[5,6], MD, PhD; Toru Horinouchi[7], MD; Ryo Kawagishi[8], MD; Marcus P J Tan[9], BSc, MRCPsych; Elizabeth R Hooker[10], MS, MD; Melchor Alvarez-Mon[5,11], MD, PhD; Alan R Teo[10,12], MD, MS

[1]Department of Child and Adolescent Psychiatry, NYU Grossman School of Medicine, New York, NY, United States

[2]Department of Psychiatry, Columbia University, New York, NY, United States

[3]Division of Translational Epidemiology, New York State Psychiatric Institute, New York, NY, United States

[4]Department of Psychiatry and Clinical Psychology, Clinica Universidad de Navarra, Pamplona, Spain

[5]Department of Medicine and Medical Specialties, Faculty of Medicine and Health Sciences, University of Alcala, Alcala de Henares, Spain

[6]Department of Psychiatry and Mental Health, Hospital Infanta Leonor, Madrid, Spain

[7]Department of Psychiatry and Neurology, Hokkaido University Graduate School of Medicine, Hokkaido, Japan

[8]Department of Psychiatry, Chiba Psychiatric Medical Center, Chiba, Japan

[9]Department of Child and Adolescent Psychiatry, South London and Maudsley NHS Foundation Trust, London, United Kingdom

[10]VA Portland Health Care System, Health Services Research & Development Center to Improve Veteran Involvement in Care, Portland, OR, United States

[11]Immune System Diseases-Rheumatology, Oncology Service and Internal Medicine, Hospital Universitario Principe de Asturias, Alcalá de Henares, Spain

[12]Department of Psychiatry, Oregon Health & Science University, Portland, OR, United States

Corresponding Author:
Miguel Angel Alvarez-Mon, MD, PhD
Department of Medicine and Medical Specialties
Faculty of Medicine and Health Sciences
University of Alcala
Campus Universitario – C/ 19, Av, de Madrid, Km 33,600
Alcala de Henares
Spain
Phone: 34 91 885 4505
Email: maalvarezdemon@icloud.com

Abstract

Background: Hikikomori is a form of severe social withdrawal that is particularly prevalent in Japan. Social media posts offer insight into public perceptions of mental health conditions and may also inform strategies to identify, engage, and support hard-to-reach patient populations such as individuals affected by hikikomori.

Objective: In this study, we seek to identify the types of content on Twitter related to hikikomori in the Japanese language and to assess Twitter users' engagement with that content.

Methods: We conducted a mixed methods analysis of a random sample of 4940 Japanese tweets from February to August 2018 using a hashtag (#hikikomori). Qualitative content analysis included examination of the text of each tweet, development of a codebook, and categorization of tweets into relevant codes. For quantitative analysis (n=4859 tweets), we used bivariate and multivariate logistic regression models, adjusted for multiple comparisons, and estimated the predicted probabilities of tweets receiving engagement (likes or retweets).

Results: Our content analysis identified 9 codes relevant to tweets about hikikomori: *personal anecdotes*, *social support*, *marketing*, *advice*, *stigma*, *educational opportunities*, *refuge (ibasho)*, *employment opportunities*, and *medicine and science*. Tweets about *personal anecdotes* were the most common (present in 2747/4859, 56.53% of the tweets), followed by *social support* (902/4859, 18.56%) and *marketing* (624/4859, 12.84%). In the adjusted models, tweets coded as *stigma* had a lower predicted

probability of likes (−33 percentage points, 95% CI −42 to −23 percentage points; *P*<.001) and retweets (−11 percentage points, 95% CI −18 to −4 percentage points; *P*<.001), *personal anecdotes* had a lower predicted probability of retweets (−8 percentage points, 95% CI −14 to −3 percentage points; *P*=.002), *marketing* had a lower predicted probability of likes (−13 percentage points, 95% CI −21 to −6 percentage points; *P*<.001), and *social support* had a higher predicted probability of retweets (+15 percentage points, 95% CI 6-24 percentage points; *P*=.001), compared with all tweets without each of these codes.

Conclusions: Japanese tweets about hikikomori reflect a unique array of topics, many of which have not been identified in prior research and vary in their likelihood of receiving engagement. Tweets often contain personal stories of hikikomori, suggesting the potential to identify individuals with hikikomori through Twitter.

(J Med Internet Res 2022;24(1):e31175) doi:10.2196/31175

KEYWORDS

hikikomori; loneliness; social isolation; social withdrawal; Twitter; hidden youth; mobile phone

Introduction

Background

Hikikomori is a form of severe social withdrawal, initially described in Japan in the 1990s, and since the 2010s, it has been increasingly reported in other countries around the globe, including the Western world [1,2]. Individuals with hikikomori are described as people who shut themselves in their homes for months and even years, with minimal interaction with society and little to no participation in school or the workforce [3]. Hikikomori can cause significant distress to the affected individuals and is often associated with psychiatric disorders [4,5]. It has also been considered a major socioeconomic and public health concern in Japan for years, with an estimated prevalence of approximately 1% [6,7].

A longstanding area of debate is whether hikikomori constitutes (or is a manifestation of) psychopathology versus sociological phenomena such as nonmainstream lifestyle preferences, cultural marginalization [8], or nonconforming reactions to societal constraints [9]. To an extent, hikikomori represents psychopathology, and additional issues include how to diagnose and treat it [7].

The nature of hikikomori makes affected individuals a hard-to-reach population [10] in terms of research and intervention efforts. Although hikikomori was described in Japan much before the *digital revolution* of the 2000s, the internet, social media, and web-based gaming have radically changed the way people interact [11]. This may be particularly relevant among individuals with hikikomori, a *hidden population* that might be spending a considerable amount of time on the internet for entertainment and social interaction [12]. Indeed, the relationships among internet use, video gaming, social media use, and hikikomori have been studied in Japan [13]. Given this, the *online world* has been proposed as an accessible gateway to reach and support individuals with hikikomori [10,14].

Social media platforms, including Twitter [15], Facebook [16], and Instagram [17] have been increasingly harnessed for health research. Twitter, a popular microblogging platform mainly based in short text posts (*tweets*), counts on >300 million users worldwide [18] and provides open access to its public contents. Health research on Twitter has included exploration of content in the public conversation regarding health conditions and treatments, engagement of users (reach of general public, recruitment of research subjects, and intervention on target populations), and real time epidemiological surveillance [15] (these applications of internet and social media-based data have been named *Infodemiology* and *infoveillance*) [19].

Twitter can be especially useful for health research in Japan, as it is the most popular social media platform in this country [20], with 51.9 million users as of October 2020 (in absolute number of Twitter users worldwide, Japan is only behind the United States, which has 68.7 million users) [18]. We previously reported findings based on analysis of tweets containing the hashtag #hikikomori [10]; the study found that tweets depicted hikikomori as either *not a problem* (eg, as a lifestyle or a nonconcerning behavior) or as a medical or social problem. Tweets with scientific content and tweets mentioning hikikomori in countries other than Japan showed significantly higher user engagement than those without these topics. However, the study was limited in sample size and only included tweets in 5 Western languages (English, Italian, Spanish, Catalan, and French).

Objectives

The objective of this study is to analyze Japanese language tweets related to hikikomori. Our two primary research questions are as follows: (1) What are the main types of content among Japanese language tweets related to hikikomori? (2) What tweets result in the most engagement (as measured by users' retweets and likes)?

Methods

Study Design and Overview

In this mixed methods study [21], we used concurrent collection and analysis of qualitative and quantitative social media data to better understand hikikomori. Qualitative data and analysis focused on content analysis of publicly available tweets about hikikomori in the Japanese language, whereas quantitative data and analysis focused on metrics of engagement with the tweets contained in the content analysis.

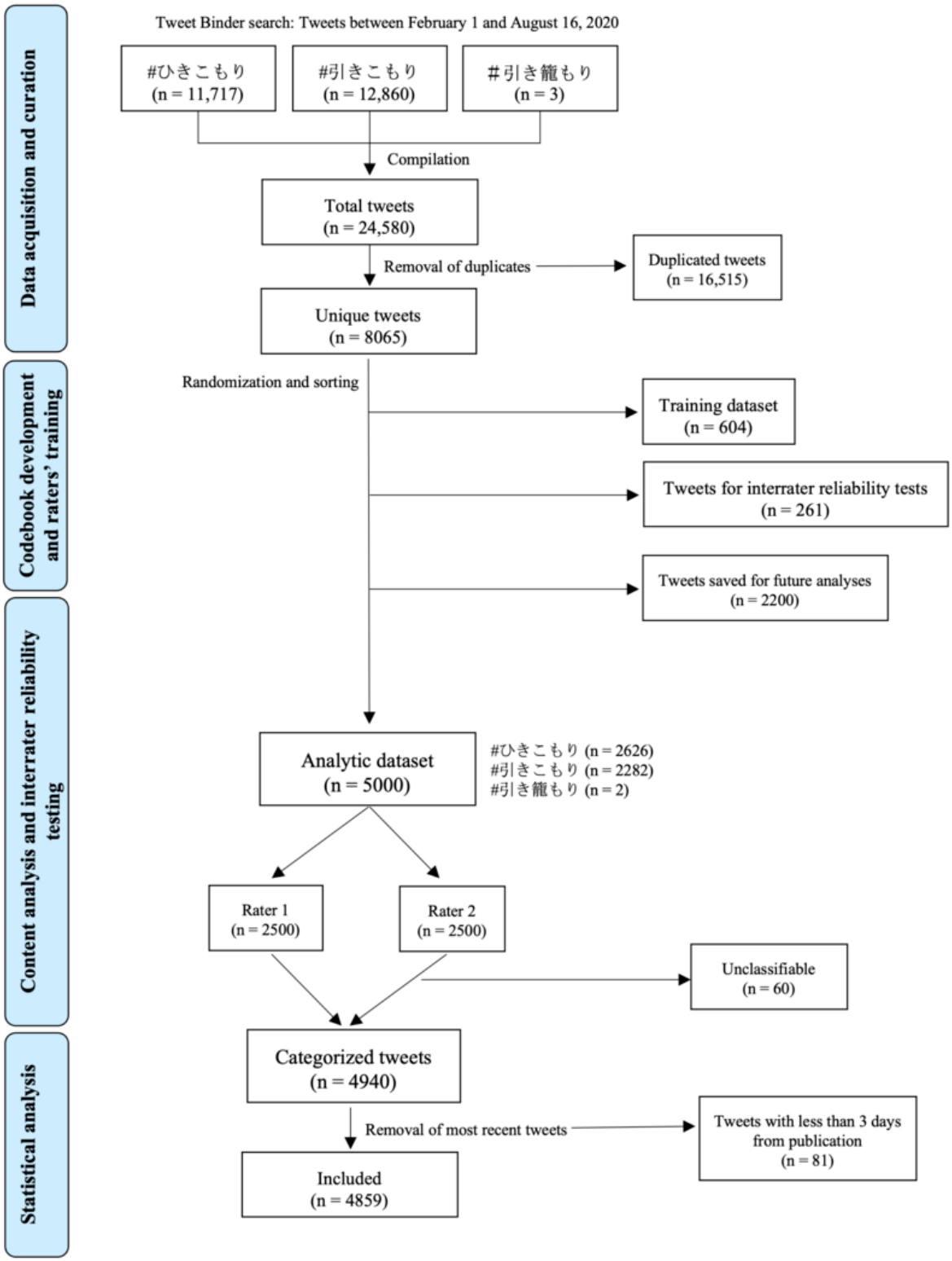
Translation (Japanese to English) was done by bilingual research team members (ART and MPJT), with backtranslation (English to Japanese) by native Japanese speakers (TH and RK). This study was approved by the University of Navarra Research Ethics Committee (ID: 2018.36-mod1) and the Veterans Affairs Portland Health System Research and Development Committee

(ID: 4524). We used publicly available tweets, which are subject to universal access according to Twitter's terms of service [22].

Data Collection and Curation

Figure 1 presents a flowchart summarizing the steps in data collection and analysis, along with the number of tweets included and excluded in each step. We used the Tweet Binder engine for the identification and collection of tweets. As described in our previous studies on Twitter content analysis, [10,23-25], Tweet Binder uses the Twitter provider Firehose via the tool Gnip (Gnip Inc) to access 100% of the public tweets matching a specific query, whereas some other search engines based on Twitter's application programing interface can only access small samples [15,26].

Figure 1. Flowchart summarizing the steps in data collection and analysis.



We included tweets that met the following criteria: (1) were public (ie, not posted as *protected* by users); (2) contained any of the 3 hashtags, each representing a way to transcribe the term *hikikomori* in Japanese (#ひきこもり or #引きこもり or #引き籠り); (3) were posted between February 1 and August 16, 2018; and (4) contained any text in Japanese besides the hashtag itself. The exclusion criteria were as follows: (1) the majority of text in the tweet was in another language besides Japanese or (2) tweets only contained a link or picture without any text. Extracted information also included metadata (date of tweets, contributors' usernames and number of followers, number of likes and retweets, frequently associated hashtags, etc). We merged the full data for each of the tweets containing 1 of the 3 hashtags into 1 data set, removed duplicates, and randomized the order of tweets in the data set.

Content Analysis: Codebook Development and Training

Training was provided by research team members (VPS and ART) experienced in content analysis and codebook development [10,16,24]. There were 2 primary coders (TH and RK), with a third research team member (MPJT) assisting with adjudication of coding disagreements.

We first created a *training data set* of 604 tweets for content exploration and for training coders. Both coders looked at the text of each tweet independently, being blind to its metadata (username and date), to identify both hypothesis-driven codes (ie, types of contents previously identified in our study on hikikomori in Western languages [10]: *hikikomori not as a problem*, *medicine and science*, *personal anecdotes*, *social*, *scientific reference*, and *hikikomori out of Japan*) and data-driven codes (new types of contents). Tweets could be coded into multiple codes when appropriate, although the assignment of a single code was preferred.

A codebook was developed using an iterative process through regular team meetings; the final codebook comprised 9 codes that fit all the main topics present in the data set and for which the interrater reliability (IRR), as measured by agreement percentages, was very high (>95%). Table 1, which lists these codes and provides definitions and examples, was used as a reference for coders in the following step.

Table 1. Codebook for content analysis (N=5000 tweets).[a]

Code	Definition	Examples of tweets
Unclassifiable	Tweets with insufficient information to be coded. These typically included brief tweets or tweets with seemingly random content with little relevance to hikikomori.	• "I'm in the intensive camp trying to get a driving license, I can't stand it anymore #event #smartphone #driverslicense #camp #Iwannagohome #hikikomori #ubearable (URL)." [Tweet ID 1695]
Personal anecdotes	Tweets describing experiences with hikikomori. These can either be from people who self-identify as hikikomori (first person stories) or comments about others thought to have hikikomori (second or third person stories).	• "A quiet morning. He[b] hasn't come into the living room. Perhaps he's having a restful sleep? 1 year and several months ago Even if he had medications he couldn't sleep. It was difficult for him to fall asleep alone. Now, even without medication, he falls asleep just like this. I hope he sleeps a lot...#depression #hikikomori #schoolrefusal." [Tweet ID 4796]
Social support	Tweets about resources that may provide social support, such as online or face-to-face support groups or hotlines for people affected by hikikomori.	• "Self-help group for those who are unemployed or on leave (URL) # Self-helpgroup "Intersection"#Unemployed #On leave #Returntowork #Depression #Hikikomori #Socialparticipation #Self-helpgroup #Createanibasho #Urawa[c]#counseling #NEET[d] #Psychiatry." [Tweet ID 1156] • "Good evening, this is Akebonobashi[c] Independence Training Center. Are you troubled by #Hikikomori #NEET #developmental disorder #domestic violence? Do discuss it with us! (URL)." [Tweet ID 1320]
Marketing	Tweets advertising or offering services to individuals with hikikomori (note: if the service being marketed was a job offer or schooling or educational opportunity, they were coded using those codes instead).	• "There aren't many people with a PC who aren't doing this you know? It's what happens when you try too hard...lol[e] ⇒ (URL) #Sidejob #Millionaire #NEET #Hikikomori." [Tweet ID 1011]
Advice	Tweet offering suggestions, recommendations, or advice for individuals with hikikomori.	• "Today, discussed "listening communication" at the "trouble with returning to work café." While communication tends to emphasize "speaking," actually "listening" can also be important! Perhaps this approach can be effective for people who have problems with communication? #support with return to work #youthsupport #ibasho #Hikikomori (URL)." [Tweet ID 172]
Stigma	Tweets using hikikomori as a pejorative word or insult.	• "Colorful, small-sized clothes are worn by both older women, and younger ladies in their 20s. Even if both of these groups have a slim and short physique, I think there are clothes that are suitable for a certain age. It's like washing everyone's clothes together on the weekend.[f] #Cigarettesmell #Noise #Annoyingbehavior #Condominiumresident #Okagami[c] #Hikikomori #Makingyoung #disgusting #Aunt." [Tweet ID 1021]
Educational opportunities	Tweets about schooling options or other educational opportunities for individuals with a diagnosis of hikikomori.	• "Kyoto or Osaka correspondence school. A school in Kyoto, which is eligible for a secure study Support System. (URL) #School refusal #Hikikomori." [Tweet ID 677]
Refuge (*Ibasho*)	Tweet that describes or offers a refuge, respite, or other safe space for people (including those affected by hikikomori). In Japanese, the term *ibasho* is used.	• "Hi there! Today at my ibasho, I had a day of art A bit like winter? Using felt to make a mini cushion! Feel like it turned out pretty good !! #Tokyo #tokyo #tokyo[g] #sugamo[c] #sugamo[c,g] #hikikomori...(URL)." [Tweet ID 399]
Employment opportunities	Tweets offering jobs for individuals with hikikomori.	• "Ocomail[h] is in production! #Starfish* #myjobistofind people #recruitment #Occupation #Kagawa[c] #Kochi[c] #Tokushima[c] #Ehime[c] #Shikoku[c] #LGBT #Hikikomori #Elderlypeople #Foreigners #Mother #Disabledpeople #Career-change #Jobhunting #Job #Staffrecruitment #Recruitment #Mid-career recruitment #Ocomail" [Tweet ID 155]
Medicine and science	Tweet related to the epidemiology, psychopathology, diagnosis, research, or treatment of hikikomori. Tweets with an explicit reference to a scientific publication, government document, or other official source are also included here.	• "20xx news from cabinet office: Hikikomori (40-64 yrs old) A national survey...'rarely leaves their own room,' 'travels only to nearby convenience stores' → counselling with Hikikomori (URL) #cabinet office #Hikikomori #nationalsurvey (URL)." [Tweet ID 11,384] • "The average age of hikikomori is 34.4 years, with an average duration of 11.8 years. The ages appear to be increasing. ⇒ The average age of actual hikikomori is 34.4 years old, with an average duration of hikikomori is 11 years 8 months. Both of which were the highest ever in an official survey. #Hikikomori (URL)." [Tweet ID 3212]

[a]This table presents the definitions and examples of the codes in our codebook. Spacing between lines and paragraphs, if present in the original tweets, was removed to shorten the length of the table. Hyperlinks, when present in the original tweet, were removed in these examples (we leave [URL] to indicate that a hyperlink was present in the original tweet). All hashtags (starting with #) were translated into English unless they used unique concepts without appropriate English equivalents.

[b]The tweet does not specify what gender the person in question is; the male pronoun is used only for the purposes of this translation.

[c]Toponym.

[d]NEET: not in employment, education, or training.

[e]The letter "W" in the original text is thought to represent an onomatopoeia for the sound of laughter. On the internet, it is used similar to its English equivalent of *lol* (ie, *laugh out loud*).

[f]It is standard practice in Japan to do one's laundry separate from others, particularly as people commonly live in shared accommodation.

[g]The Japanese language comprises of multiple writing systems; here, the same toponyms are spelled in different hashtags using different writing systems.

[h]*Ocomail* is a popular Japanese company that specializes in shipping locally grown Japanese rice overseas.

Content Analysis: Categorization of Tweets

An independent subsample of 5000 tweets (*analytic data set*) was used for content analysis, and the newly developed codebook was applied. Each coder independently examined the texts of 50% (2500/5000) tweets. For each tweet, raters were instructed to determine whether it fit the inclusion criteria (60/5000, 1.2% tweets were *unclassifiable* and excluded) and code it for the presence or absence of each of the 9 codes.

To ensure acceptable IRR and prevent *coder drift* [27], both raters also coded 3 batches of tweets from an independent subset (*IRR data set*), 1 batch per week during the first 3 weeks of the content analyses. The lead investigator (VPS) monitored the IRR for these batches and provided interim feedback when appropriate.

Statistical Analyses: IRR and Engagement Metrics

Statistical analyses were conducted using Stata 16 (StataCorp) and included calculations of IRR, descriptive figures of the distribution of tweets by codes, and analysis of engagement metrics.

We calculated the agreement percentages to assess the IRR for the 261 double-coded tweets from the *IRR data set*, which were used for coder training and not used during content analyses. Agreement percentages (presented as last batch of double coding/average across batches of double coding) were 79.21%/77.84% for *personal anecdotes*, 91.09%/90.68% for *social support*, 87.13%/85.71% for *marketing*, 91.09%/90.68% for *advice*, 92.08%/89.44% for *stigma*, 98.02%/96.89% for *educational opportunities*, 98.02%/96.89% for *refuge* (*ibasho*), 99.01%/98.76% for *employment opportunities*, and 99.01%/99.38% for *medicine and science*. We use agreement percentages over κ coefficients, as the latter underestimates agreement when the *prevalence* values (in our case, number of tweets) for a category or code are too low [27].

As for users' engagement metrics, we analyzed likes and retweets for each code in the analytic data set. Previous research has estimated that most users' engagement with social media content occurs within a week after posting [28]. Therefore, we further restricted our analytic sample to tweets that had at least 3 days of follow-up between posting and data collection, thereby excluding 1.64% (81/4940) tweets.

Bivariate logistic regression models were used to test for the association between tweets and the presence of a code and receiving at least one like or retweet. Multivariate models tested the same association with adjustment for (1) user's number of tweets in our analytic sample, (2) user's number of followers, and (3) number of days between posting and data collection. All models were clustered by the user. Results were presented as proportion points difference (also understood as difference in predicted probabilities between tweets with and without the code) rather than model coefficients for ease of interpretation [29]. Critical values for Bonferroni correction for multiple comparisons were calculated by dividing the α level (.05) by the number of hypotheses (9) and applied to all results (critical value $P<.005$). These regression models were preferred over linear regressions for numbers of likes and retweets because of the uneven distribution of these numbers among tweets.

Results

As detailed in the flowchart presented in Figure 1, of the 8065 tweets collected by the Tweet Binder tool, 4940 (61.25%) and 4859 (60.25%) unique tweets were included in the qualitative content and quantitative data analysis, respectively. A total of 1680 unique users contributed to those tweets, with an average of 1 tweet per user (median 1; IQR 1-2); only 54 users (3.21%) contributed >10 tweets.

Content of Tweets That Reference Hikikomori

The codebook applied to the analytic data set included 9 codes, 1 of which was hypothesis-driven from our previous research [10] and the remaining 8 were data-driven based on the exploration of tweets in the *training data set*:

1. Hypothesis-driven code: *Medicine and science* included tweets related to epidemiological, therapeutic, or research aspects of hikikomori understood as a pathology (eg, a tweet about a published research paper on hikikomori).
2. Data-driven codes: *Marketing*, *employment opportunities*, and *educational opportunities* included different kinds of offers apparently targeted at people with hikikomori. *Social support* and *refuge* (*ibasho*) comprised tweets promoting resources to help people with hikikomori, such as online or onsite *social support* groups, hotlines, or *ibasho*—a Japanese concept referring to designated spaces of psychological comfort for people in distress [30]. *Personal anecdotes* were related to stories of people describing hikikomori symptoms or behaviors with or without negative connotations (*stigma*) or helpful information (*advice*).

As mentioned earlier, definitions and examples are available in Table 1.

The 10 most frequently used hashtags are represented in Figure 2. The most common hashtags were related to education and employment (#school absenteeism or #refusal and #NEET, an acronym used to refer to people *not in education, employment, nor training*), whereas some less frequent ones were related to mental health and support.

Figure 2. Word cloud illustrating the 10 most frequently used hashtags in the tweets analyzed.

Japanese word	Count	Contextualized translation
不登校	907	School absenteeism or refusal
ニート	759	Not in education, employment nor training (NEET)
副業	99	Side or second job
億万長者	93	Millionaire
転職	74	Career change
無職	63	Unemployed
心理カウンセラー	60	Therapist/Counselor
鬱	31	Depression
休職中	12	On leave (from work)
自助グループ	8	Peer support (“self-help”) group

Distribution of Tweets by Codes

The tweet contents were unevenly distributed across the codes. Of the 4859 tweets, the code *personal anecdotes* was present in 2747 (56.53%), whereas *social support* was present in 902 (18.56%) tweets, and *marketing* was present in 624 (12.84%) tweets. To note, *medicine and science* was the code with the least tweets (31/4859, 0.63%). Complete figures on the number and percentages of tweets per code are presented in Table 2, along with the descriptive figures for likes and retweets.

Table 2. Descriptive characteristics of the tweets included in the analysis by code (N=4859).[a]

Code	Tweets, n (%)	Likes		Retweets	
		At least one[b], n (%)	Median[c] (IQR)	At least one[b], n (%)	Median[c] (IQR)
Personal anecdotes	2747 (56.53)	1318 (48)	3 (1-7)	436 (15.9)	1 (1-3)
Social support	902 (18.56)	276 (30.6)	2 (1-4)	211 (23.4)	1 (1-3)
Marketing	624 (12.84)	199 (31.9)	2 (1-5)	106 (17)	1 (1-2)
Advice	281 (5.78)	93 (33.1)	2 (1-5)	60 (21.4)	1 (1-3)
Stigma	166 (3.42)	12 (7.2)	1 (1-3)	9 (5.4)	1 (1-1)
Educational opportunities	129 (2.65)	37 (28.7)	2 (1-5)	21 (16.3)	2 (1-3)
Refuge (*Ibasho*)	86 (1.77)	40 (46.5)	4 (1-7)	30 (34.9)	2 (1-4)
Employment opportunities	82 (1.69)	28 (34.2)	3 (1-13)	21 (25.6)	3 (1-15)
Medicine and science	31 (0.64)	16 (51.6)	2 (1-3)	7 (22.6)	4 (2-7)

[a]For each code, the total number of tweets and retweets (n) and relative proportions (%) are provided. The total number of tweets in the first column may add to more than the total number of tweets that we have analyzed because 1 tweet could be coded into multiple codes.

[b]Among tweets in the code, n (%) with at least one like (or retweet).

[c]Among tweets in the code which had at least one like (or retweet), median (IQR) of the number of likes (or retweets).

Engagement Metrics by Codes

Approximately half of all tweets in *medicine and science*, *personal anecdotes*, and *refuge* (*ibasho*) codes received at least one like (16/31, 51.6%; 1318/2747, 48%; and 40/86, 46.5%, respectively), whereas one-third (30/86, 35%) of tweets in *refuge* (*ibasho*)' and approximately one-quarter of tweets in *employment opportunities* (21/82, 26%) and *social support* (211/902, 23.4%) received at least one retweet. Tweets with the code *stigma* had the lowest probability of having at least one like (12/166, 7.2%) or 1 retweet (9/166, 5.4%; Table 2).

Results from logistic regression analyses with clustering by user are presented in Table 3 (for likes) and Table 4 (for retweets). In unadjusted models, tweets with the *stigma* code had a significantly lower predicted probability of receiving likes (−35 percentage points, 95% CI −45 to −25 percentage points; $P<.001$) and receiving retweets (−13 percentage points, 95% CI −20 to −7 percentage points; $P<.001$) compared with all tweets without that code. Tweets coded as *personal anecdotes* had a significantly higher predicted probability of receiving likes (+16 percentage points, 95% CI 3-29 percentage points; $P=.02$) compared with all tweets without that code. No other associations between codes and being liked or retweeted were significant in the unadjusted models.

Table 3. Association between content analysis codes and receiving at least one like (N=4859 tweets) using logistic regression with adjustment for covariates and clustering by user.[a]

Code	Estimated probability (95% CI) by tweet content			*P* value
	Tweets without code (%)	Tweets with code (%)	Difference (percentage points)	
Personal anecdotes	35.4 (30.3 to 40.5)	44.9 (36.5 to 53.3)	9.5 (0.5 to 18.5)	.04[b]
Social support	41 (34.1 to 47.9)	41.2 (32.9 to 49.4)	0.2 (−12.6 to 12.9)	.98
Marketing	42.9 (37.6 to 48.1)	29.5 (19.8 to 39.3)	−13.3 (−20.8 to −5.9)	<.001[b,c]
Advice	41.1 (35.2 to 47.1)	39.7 (26.3 to 53.1)	−1.4 (−16.9 to 14.0)	.86
Stigma	42.1 (36.4 to 47.7)	9.5 (2.6 to 16.3)	−32.6 (−41.9 to −23.3)	<.001[b,c]
Educational opportunities	41.5 (35.8 to 47.1)	27.5 (16.7 to 38.4)	−13.9 (−25.4 to −2.4)	.02[b,c]
Refuge (*Ibasho*)	41.1 (35.5 to 46.6)	40.5 (16.4 to 64.7)	−0.5 (−24.5 to 23.5)	.97
Employment opportunities	41.2 (35.6 to 46.8)	33.7 (21.9 to 45.4)	−7.5 (−20.2 to 5.1)	.24
Medicine and science	41 (35.5 to 46.6)	42.4 (23.7 to 61)	1.3 (−17.5 to 20.2)	.89

[a]Results are expressed as the difference in predicted probability of at least one like between tweets with and without the code, wherein a positive value indicates a higher probability of receiving a like among tweets with the code present compared with tweets without the code. Models adjusted for (1) number of user tweets in the data set, (2) number of followers for the user, and (3) number of days between posting and the data collection date.

[b]Significance at critical value $P<.05$.

[c]Significant at the Bonferroni-adjusted critical value $P<.006$.

Table 4. Association between content analysis code and receiving at least one retweet (N=4859 tweets) using logistic regression with adjustment for covariates and clustering by user.[a]

Code	Estimated probability (95% CI) by tweet content			*P* value
	Tweets without code (%)	Tweets with code (%)	Difference (percentage points)	
Personal anecdotes	23.2 (19.1 to 27.3)	14.9 (11.2 to 18.6)	−8.3 (−13.6 to −3.1)	.002[b,c]
Social support	16.1 (12.9 to 19.3)	31.4 (23.6 to 39.3)	15.3 (6.3 to 24.3)	.001[b,c]
Marketing	18.7 (15.6 to 21.8)	15.2 (10 to 20.3)	−3.5 (−8.2 to 1.2)	.14
Advice	17.8 (14.6 to 21)	25.6 (18.7 to 32.5)	7.8 (−0.3 to 16.0)	.06
Stigma	18.5 (15.4 to 21.6)	7.5 (1.7 to 13.4)	−11.0 (−17.6 to −4.3)	.001[b,c]
Educational opportunities	18.3 (15.2 to 21.3)	15.7 (6.9 to 24.5)	−2.5 (−11.5 to 6.4)	.58
Refuge (*Ibasho*)	18 (15 to 20.9)	29 (8.8 to 4.9)	11 (−8.7 to 30.7)	.27
Employment opportunities	18.1 (15 to 21.1)	24.8 (11.1 to 38.5)	6.8 (−7.1 to 20.6)	.34
Medicine and science	18.2 (15.1 to 21.2)	17.1 (3.8 to 30.4)	−1.1 (−14.4 to 12.2)	.87

[a]Results are expressed as the difference in predicted probability of at least one retweet between tweets with and without the code, where a positive value indicates a higher probability of receiving a retweet among tweets with the code present compared with tweets without the code. Models adjusted for (1) number of user tweets in the data set, (2) number of followers for the user, and (3) number of days between posting and data collection date.

[b]Significance at critical value $P<.05$.

[c]Significant at the Bonferroni-adjusted critical value $P<.006$.

In adjusted models, the associations between *stigma* and lower predicted probability of being liked (−33 percentage points, 95% CI −42 to −23 percentage points; $P<.001$) and retweeted (−11 percentage points, 95% CI −18 to −4 percentage points; $P<.001$) remained highly significant, whereas the association of *personal anecdotes* with a higher predicted probability of being liked lost significance. In contrast, several associations that were not significant in the unadjusted models became significant after adjustment. Tweets in *marketing* had a significantly lower predicted probability of receiving likes (−13 percentage points, 95% CI −21 to −6 percentage points; $P<.001$), tweets with *personal anecdotes* had a significantly lower predicted probability of receiving retweets (−8 percentage points, 95% CI −14.0 to −3 percentage points; $P=.002$), and tweets with *social support* had a significantly higher probability of receiving retweets (+15 percentage points, 95% CI 6-24 percentage points; $P=.001$), compared with all tweets without each of these codes. These associations were statistically significant at the false discovery rate critical value ($P<.017$) to account for multiple comparisons. Other associations that were significant at the level of $P<.05$ are presented in Tables 3 and 4.

Discussion

Principal Findings

Our mixed method analysis of nearly 5000 Japanese language tweets revealed a unique array of topics discussed in relation to hikikomori, many of which have not been identified in prior studies. Personal anecdotes about hikikomori predominated, suggesting that individual Twitter users are willing to share their personal stories and experiences with hikikomori on social media. School absenteeism (*futoko*) and withdrawal from the education system and labor force (*not in employment, education, or training*) were also commonly associated with the hikikomori hashtag, adding corroboration from Twitter data that these 2 concepts are closely linked to the lives of people with hikikomori and frequently discussed in Japan [8,31]. Engagement (retweets and likes) varied by tweet content, but tweets with stigmatizing content received consistently lower engagement.

To the best of our knowledge, this report presents the first application of social media research to a data set of Japanese tweets related to hikikomori. This study builds upon our previous study of tweets with #hikikomori in several Western languages [10], both by examining a significantly larger data set and also by identifying a distinct set of topics within the Japanese Twitter discourse on hikikomori. In contrast to our prior study, this study revealed that tweets in Japanese tend to relate to personal stories (*personal anecdotes*) of hikikomori, as well as *marketing* (in many cases, presented as *click-bait* [32]) and social support opportunities targeting individuals with hikikomori. These findings support the suspicion that social media may indeed be a refuge for individuals with hikikomori and serve as a place where they can find social support [10,14].

It is noteworthy that the code *medicine and science* was by far the least identified in the Japanese data set, accounting for <1% of the tweets, in contrast to our previous study on tweets in Western languages, where these contents were present in 42.22% of classifiable tweets [10]. This suggests the existence of cross-cultural differences in the way hikikomori is conceived and discussed by the general public in Japan versus Western countries. Although hikikomori seems to be a term more integrated in popular culture and a part of one's identity in Japan, Western countries tend to view it as a worrisome behavior and related to mental health issues.

Our discovery of stigmatizing tweets eliciting *negative* public engagement (lower predicted probability of retweets and likes) is worth discussing. Stigma, a social phenomenon involving

negative attitudes toward people with certain characteristics or conditions, markedly affects people with mental disorders [33]. Given the potential role that social media plays in the perpetuation of misinformation, stereotypes, and hateful speech, psychiatric research in this area has particularly focused on stigma [34]; examples include studies on psychosis and schizophrenia [24,35,36], bipolar disorder [37], and the depiction of mental disorders by mass media [38]. The infrequency of stigmatizing Twitter content related to hikikomori and the relative lack of engagement with such content are hopeful findings.

A final point for discussion of our results relates to our research question about the differential patterns of public engagement (retweets and likes) generated by each of the topics. Engagement with social media content, apart from being a marker of visibility, may reflect the public's interest, perceptions, and behavior [23,28,39]. One possible explanation for the pattern of a higher probability of likes but a lower predicted probability for retweets observed for *personal anecdotes* tweets is that Twitter users may show solidarity with the person disclosing hikikomori but less willingness to publicly share and endorse those personal stories with their own followers.

Our social media research on hikikomori, the results of which are presented in our previous study of Western language tweets [10] and this study, constitutes the first application of Twitter content analyses to this phenomenon. To contextualize our study in the scientific literature, 2 aspects are worth noting. On the one hand, our methods were built on previous social media studies in the area of health [24,25,28,37] and incorporated innovations based on our own hypotheses and the retrieved data. Social media research is relatively young, and the preferred methodology is subject to change. In contrast, recent hikikomori research has paid more attention to the interplay among social withdrawal, smartphones and technology, internet use and addiction, and social media, where causal relationships seem difficult to untangle [11,13]. Further work is needed to fine-tune and replicate Twitter analysis methods in health research, as well as to study social media use by people with hikikomori, examining the patterns of use, contents they consume and generate and their influence on them and on the general public, and avenues for research and public health interventions to reach and support them.

Limitations

The main limitations of this study are as follows: (1) hyperlinks that were included in the original tweets were not analyzed, which limited the ability to understand the full context of the tweets; (2) other tweets potentially related to hikikomori may have been missed if they used other hashtags not captured in our data set; (3) although high overall, the IRR was variable across codes and weaker in some of them, especially in *personal anecdotes* code; (4) metrics of engagement (likes and retweets) may have been influenced by unknown or unmeasurable confounding factors (eg, characteristics of the user posting the content, factors related to the user's followers, and other contextual factors).

Conclusions

In conclusion, Japanese tweets that are related to hikikomori are abundant and contain a wide array of topics. Engagement patterns varied but stigmatizing and marketing content were generally less likely to receive engagement, whereas personal stories and social support showed some evidence of being more likely to receive engagement. Future research to better understand the characteristics that make some tweets more likely to elicit reactions [40,41], their significance [42], and the intriguing ways in which retweets and likes converge and diverge [24] would be helpful. Our findings can inform Twitter content to potentially identify and connect with this hard-to-reach population.

Acknowledgments

ART's work was supported in part by a Career Development Award from the Veterans Health Administration Health Service Research and Development (CDA 14-428). The US Department of Veterans Affairs had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication. The findings and conclusions in this document are those of the authors who are responsible for its contents; the findings and conclusions do not necessarily represent the views of the US Department of Veterans Affairs or the Government of United States. The authors would like to acknowledge the Japanese Society of Psychiatry and Neurology for the past Fellowship Awards granted to ART, VPS, and MPJT and for their encouragement to work on international research studies on hikikomori. Ms Teresa Abrego and Ms Maite Muruzabal from Tweet Binder, Spain collaborated significantly in the retrieval of tweets through their search engine. The authors would also like to thank Justin S Yin for proofreading the manuscript.

This work was partially supported by grants from the *Fondo de Investigación de la Seguridad Social*, Instituto de Salud Carlos III (PI18/01726), Spain and the *Programa de Actividades de I+D de la Comunidad de Madrid en Biomedicina* (B2017/BMD-3804), Madrid, Spain.

Authors' Contributions

VPS and MAAM were the principal contributors for the research design, coordination of data analysis, and manuscript preparation; VPS specifically coordinated the rater's training and interrater reliability assessment and discussion; MAAM specifically coordinated the data acquisition. TH and RK were the Japanese tweet coders, contributing to the codebook development, training,

and analysis of the tweets. MPJT contributed mainly to the Japanese–English interpreter and mediator during raters' training and interrater reliability discussions, as well as in the development of the codebook. ERH conducted and reported statistical analyses. MAM contributed as a reviewer of the manuscript. ART was the main supervisor of all phases of the project, with special involvement in the study design, interpretation of data, and manuscript preparation.

Conflicts of Interest

None declared.

References

1. Malagón-Amor A, Martín-López LM, Córcoles D, González A, Bellsolà M, Teo AR, et al. A 12-month study of the hikikomori syndrome of social withdrawal: clinical characterization and different subtypes proposal. Psychiatry Res 2018 Dec;270:1039-1046. [doi: 10.1016/j.psychres.2018.03.060] [Medline: 29615267]
2. Frankova I. Similar but different: psychological and psychopathological features of primary and secondary hikikomori. Front Psychiatry 2019;10:558 [FREE Full text] [doi: 10.3389/fpsyt.2019.00558] [Medline: 31447713]
3. Kato TA, Kanba S, Teo AR. A 39-year-old "Adultolescent": understanding social withdrawal in Japan. Am J Psychiatry 2016 Feb 01;173(2):112-114 [FREE Full text] [doi: 10.1176/appi.ajp.2015.15081034] [Medline: 26844793]
4. Teo AR, Gaw AC. Hikikomori, a Japanese culture-bound syndrome of social withdrawal?: a proposal for DSM-5. J Nerv Ment Dis 2010 Jun;198(6):444-449 [FREE Full text] [doi: 10.1097/NMD.0b013e3181e086b1] [Medline: 20531124]
5. Teo AR, Stufflebam K, Saha S, Fetters MD, Tateno M, Kanba S, et al. Psychopathology associated with social withdrawal: idiopathic and comorbid presentations. Psychiatry Res 2015 Jul 30;228(1):182-183. [doi: 10.1016/j.psychres.2015.04.033] [Medline: 25977071]
6. Teo AR. A new form of social withdrawal in Japan: a review of hikikomori. Int J Soc Psychiatry 2010 Mar;56(2):178-185 [FREE Full text] [doi: 10.1177/0020764008100629] [Medline: 19567455]
7. Kato TA, Kanba S, Teo AR. Hikikomori : multidimensional understanding, assessment, and future international perspectives. Psychiatry Clin Neurosci 2019 Aug;73(8):427-440 [FREE Full text] [doi: 10.1111/pcn.12895] [Medline: 31148350]
8. Uchida Y, Norasakkunkit V. The NEET and hikikomori spectrum: assessing the risks and consequences of becoming culturally marginalized. Front Psychol 2015;6:1117 [FREE Full text] [doi: 10.3389/fpsyg.2015.01117] [Medline: 26347667]
9. Toivonen T, Norasakkunkit V, Uchida Y. Unable to conform, unwilling to rebel? Youth, culture, and motivation in globalizing Japan. Front Psychol 2011;2:207 [FREE Full text] [doi: 10.3389/fpsyg.2011.00207] [Medline: 21949510]
10. Pereira-Sanchez V, Alvarez-Mon MA, Barco AA, Alvarez-Mon M, Teo A. Exploring the extent of the hikikomori phenomenon on Twitter: mixed methods study of western language tweets. J Med Internet Res 2019 May 29;21(5):e14167 [FREE Full text] [doi: 10.2196/14167] [Medline: 31144665]
11. Kato TA, Shinfuku N, Tateno M. Internet society, internet addiction, and pathological social withdrawal: the chicken and egg dilemma for internet addiction and hikikomori. Curr Opin Psychiatry 2020 May;33(3):264-270. [doi: 10.1097/YCO.0000000000000601] [Medline: 32101902]
12. Wong M. Hidden youth? A new perspective on the sociality of young people 'withdrawn' in the bedroom in a digital age. New Media Society 2020 Jul 22;22(7):1227-1244. [doi: 10.1177/1461444820912530]
13. Tateno M, Teo AR, Ukai W, Kanazawa J, Katsuki R, Kubo H, et al. Internet addiction, smartphone addiction, and hikikomori trait in japanese young adult: social isolation and social network. Front Psychiatry 2019 Jul 10;10:455 [FREE Full text] [doi: 10.3389/fpsyt.2019.00455] [Medline: 31354537]
14. Liu LL, Li TM, Teo AR, Kato TA, Wong PW. Harnessing social media to explore youth social withdrawal in three major cities in China: cross-sectional web survey. JMIR Ment Health 2018 May 10;5(2):e34 [FREE Full text] [doi: 10.2196/mental.8509] [Medline: 29748164]
15. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a tool for health research: a systematic review. Am J Public Health 2017 Dec;107(1):1-8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
16. Franz D, Marsh HE, Chen JI, Teo AR. Using Facebook for qualitative research: a brief primer. J Med Internet Res 2019 Aug 13;21(8):e13544 [FREE Full text] [doi: 10.2196/13544] [Medline: 31411143]
17. Fung IC, Blankenship EB, Ahweyevu JO, Cooper LK, Duke CH, Carswell SL, et al. Public health implications of image-based social media: a systematic review of Instagram, Pinterest, Tumblr, and Flickr. Perm J 2020;24 [FREE Full text] [doi: 10.7812/TPP/18.307] [Medline: 31852039]
18. Twitter: most users by country. Statista. URL: https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/ [accessed 2020-12-02]
19. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
20. Social media stats Japan. StatCounter Glob Stats. URL: https://gs.statcounter.com/social-media-stats/all/japan [accessed 2021-12-02]

21. Creswell J, Plano CV, Gutmann M, Hanson W. Advanced mixed methods research designs. In: Tashakkori A, Teddlie C, editors. Handbook of Mixed Methods in Social and Behavioural Research. Thousand Oaks, CA: SAGE Publications; 2003:209-240.
22. Twitter Terms of Service. URL: https://twitter.com/en/tos [accessed 2021-12-20]
23. Alvarez-Mon MA, Barco AA, Lahera G, Quintero J, Ferre F, Pereira-Sanchez V, et al. Increasing interest of mass communication media and the general public in the distribution of tweets about mental disorders: observational study. J Med Internet Res 2018 May 28;20(5):e205 [FREE Full text] [doi: 10.2196/jmir.9582] [Medline: 29807880]
24. Alvarez-Mon MA, Llavero-Valero M, Sánchez-Bayona R, Pereira-Sanchez V, Vallejo-Valdivielso M, Monserrat J, et al. Areas of interest and stigmatic attitudes of the general public in five relevant medical conditions: thematic and quantitative analysis using Twitter. J Med Internet Res 2019 May 28;21(5):e14110 [FREE Full text] [doi: 10.2196/14110] [Medline: 31140438]
25. Viguria I, Alvarez-Mon MA, Llavero-Valero M, Barco AA, Ortuño F, Alvarez-Mon M. Eating disorder awareness campaigns: thematic and quantitative analysis using Twitter. J Med Internet Res 2020 Jul 14;22(7):e17626 [FREE Full text] [doi: 10.2196/17626] [Medline: 32673225]
26. Morstatter F, Pfeffer J, Liu H, Carley K. Is the sample good enough? Comparing data from Twitter's streaming API with Twitter's Firehose. arXiv. 2013. URL: http://arxiv.org/abs/1306.5204 [accessed 2021-12-20]
27. Neuendorf KA. Chapter 6. Reliability. In: The Content Analysis Guidebook. Thousand Oaks, CA: SAGE Publications, Inc; 2017.
28. Edney S, Bogomolova S, Ryan J, Olds T, Sanders I, Maher C. Creating engaging health promotion campaigns on social media: observations and lessons from Fitbit and Garmin. J Med Internet Res 2018 Dec 10;20(12):e10911 [FREE Full text] [doi: 10.2196/10911] [Medline: 30530449]
29. Williams R. Using the margins command to estimate and interpret adjusted predictions and marginal effects. Stata J 2012 Jun 01;12(2):308-331. [doi: 10.1177/1536867x1201200209]
30. Kunikata H, Shiraishi Y, Nakajima K, Tanioka T, Tomotake M. The relationship between psychological comfort space and self-esteem in people with mental disorders. J Med Invest 2011 Feb;58(1-2):56-62 [FREE Full text] [doi: 10.2152/jmi.58.56] [Medline: 21372487]
31. Tan MP, Lee W, Kato TA. International experience of hikikomori (prolonged social withdrawal) and its relevance to psychiatric research. BJPsych Int 2021 May 11;18(2):34-37 [FREE Full text] [doi: 10.1192/bji.2020.20] [Medline: 34287404]
32. The science behind clickbait titles (and how to use them responsibly). Brafton. 2020. URL: https://www.brafton.com/blog/creation/clickbait-titles/ [accessed 2021-05-14]
33. Sartorius N. Stigma and mental health. Lancet 2007 Sep;370(9590):810-811. [doi: 10.1016/s0140-6736(07)61245-8]
34. Robinson P, Turk D, Jilka S, Cella M. Measuring attitudes towards mental health using social media: investigating stigma and trivialisation. Soc Psychiatry Psychiatr Epidemiol 2019 Jan;54(1):51-58. [doi: 10.1007/s00127-018-1571-5] [Medline: 30069754]
35. Joseph AJ, Tandon N, Yang LH, Duckworth K, Torous J, Seidman LJ, et al. #Schizophrenia: use and misuse on Twitter. Schizophr Res 2015 Jul;165(2-3):111-115. [doi: 10.1016/j.schres.2015.04.009] [Medline: 25937459]
36. Passerello GL, Hazelwood JE, Lawrie S. Using Twitter to assess attitudes to schizophrenia and psychosis. BJPsych Bull 2019 Aug;43(4):158-166 [FREE Full text] [doi: 10.1192/bjb.2018.115] [Medline: 30784393]
37. Budenz A, Klassen A, Purtle J, Yom Tov E, Yudell M, Massey P. Mental illness and bipolar disorder on Twitter: implications for stigma and social support. J Ment Health 2020 Apr;29(2):191-199. [doi: 10.1080/09638237.2019.1677878] [Medline: 31694433]
38. Bowen M, Lovell A. Stigma: the representation of mental health in UK newspaper Twitter feeds. J Ment Health 2019 May 10:1-7. [doi: 10.1080/09638237.2019.1608937] [Medline: 31072175]
39. Dragseth MR. Building student engagement through social media. J Polit Sci Edu 2019 Feb 04;16(2):243-256. [doi: 10.1080/15512169.2018.1550421]
40. Suh B, Hong L, Pirolli P, Chi E. Want to be retweeted? Large scale analytics on factors impacting retweet in Twitter network. In: Proceedings of the IEEE Second International Conference on Social Computing. 2010 Presented at: IEEE Second International Conference on Social Computing; Aug. 20-22, 2010; Minneapolis, MN, USA p. 177-184. [doi: 10.1109/socialcom.2010.33]
41. Yang Q, Tufts C, Ungar L, Guntuku S, Merchant R. To retweet or not to retweet: understanding what features of cardiovascular tweets influence their retransmission. J Health Commun 2018;23(12):1026-1035 [FREE Full text] [doi: 10.1080/10810730.2018.1540671] [Medline: 30404564]
42. Majmundar A, Allem J, Boley Cruz T, Unger JB. The Why We Retweet scale. PLoS One 2018 Oct 18;13(10):e0206076 [FREE Full text] [doi: 10.1371/journal.pone.0206076] [Medline: 30335827]

Abbreviations

IRR: interrater reliability

Edited by R Kukafka; submitted 11.06.21; peer-reviewed by I Alberdi-Páramo, K Patel, K Wall, L Hong, D Huang; comments to author 29.08.21; revised version received 22.10.21; accepted 29.10.21; published 11.01.22.

Please cite as:
Pereira-Sanchez V, Alvarez-Mon MA, Horinouchi T, Kawagishi R, Tan MPJ, Hooker ER, Alvarez-Mon M, Teo AR
Examining Tweet Content and Engagement of Users With Tweets About Hikikomori in Japanese: Mixed Methods Study of Social Withdrawal
J Med Internet Res 2022;24(1):e31175
URL: https://www.jmir.org/2022/1/e31175
doi:10.2196/31175
PMID:35014971

Original Paper

Medical and Health-Related Misinformation on Social Media: Bibliometric Study of the Scientific Literature

Andy Wai Kan Yeung[1,2], PhD; Anela Tosevska[2,3], PhD; Elisabeth Klager[2], MSc; Fabian Eibensteiner[2,4], MD; Christos Tsagkaris[5], MD; Emil D Parvanov[2,6], PhD; Faisal A Nawaz[7], MBBSc; Sabine Völkl-Kernstock[2], PhD; Eva Schaden[2,8], MD; Maria Kletecka-Pulker[2], PhD; Harald Willschke[2,8], MD; Atanas G Atanasov[2,9], PhD

[1]Oral and Maxillofacial Radiology, Applied Oral Sciences and Community Dental Care, Faculty of Dentistry, The University of Hong Kong, Hong Kong, China

[2]Ludwig Boltzmann Institute for Digital Health and Patient Safety, Medical University of Vienna, Vienna, Austria

[3]Department of Molecular, Cell and Developmental Biology, University of California Los Angeles, Los Angeles, CA, United States

[4]Division of Pediatric Nephrology and Gastroenterology, Comprehensive Center for Pediatrics, Medical University of Vienna, Vienna, Austria

[5]Faculty of Medicine, University of Crete, Heraklion, Greece

[6]Department of Translational Stem Cell Biology, Medical University of Varna, Varna, Bulgaria

[7]College of Medicine, Mohammed Bin Rashid University of Medicine and Health Sciences, Dubai, United Arab Emirates

[8]Department of Anaesthesia, Intensive Care Medicine and Pain Medicine, Medical University of Vienna, Vienna, Austria

[9]Institute of Genetics and Animal Biotechnology of the Polish Academy of Sciences, Jastrzebiec, Poland

Corresponding Author:
Atanas G Atanasov, PhD
Ludwig Boltzmann Institute for Digital Health and Patient Safety
Medical University of Vienna
Spitalgasse 23
Vienna, 1090
Austria
Phone: 43 664 1929 852
Email: Atanas.Atanasov@dhps.lbg.ac.at

Abstract

Background: Social media has been extensively used for the communication of health-related information and consecutively for the potential spread of medical misinformation. Conventional systematic reviews have been published on this topic to identify original articles and to summarize their methodological approaches and themes. A bibliometric study could complement their findings, for instance, by evaluating the geographical distribution of the publications and determining if they were well cited and disseminated in high-impact journals.

Objective: The aim of this study was to perform a bibliometric analysis of the current literature to discover the prevalent trends and topics related to medical misinformation on social media.

Methods: The Web of Science Core Collection electronic database was accessed to identify relevant papers with the following search string: ALL=(misinformati* OR “wrong informati*” OR disinformati* OR “misleading informati*” OR “fake news*”) AND ALL=(medic* OR illness* OR disease* OR health* OR pharma* OR drug* OR therap*) AND ALL=(“social media*” OR Facebook* OR Twitter* OR Instagram* OR YouTube* OR Weibo* OR Whatsapp* OR Reddit* OR TikTok* OR WeChat*). Full records were exported to a bibliometric software, VOSviewer, to link bibliographic information with citation data. Term and keyword maps were created to illustrate recurring terms and keywords.

Results: Based on an analysis of 529 papers on medical and health-related misinformation on social media, we found that the most popularly investigated social media platforms were Twitter (n=90), YouTube (n=67), and Facebook (n=57). Articles targeting these 3 platforms had higher citations per paper (>13.7) than articles covering other social media platforms (Instagram, Weibo, WhatsApp, Reddit, and WeChat; citations per paper <8.7). Moreover, social media platform–specific papers accounted for 44.1% (233/529) of all identified publications. Investigations on these platforms had different foci. Twitter-based research explored cyberchondria and hypochondriasis, YouTube-based research explored tobacco smoking, and Facebook-based research studied vaccine hesitancy related to autism. COVID-19 was a common topic investigated across all platforms. Overall, the United States contributed to half of all identified papers, and 80% of the top 10 most productive institutions were based in this country. The

identified papers were mostly published in journals of the categories public environmental and occupational health, communication, health care sciences services, medical informatics, and medicine general internal, with the top journal being the Journal of Medical Internet Research.

Conclusions: There is a significant platform-specific topic preference for social media investigations on medical misinformation. With a large population of internet users from China, it may be reasonably expected that Weibo, WeChat, and TikTok (and its Chinese version Douyin) would be more investigated in future studies. Currently, these platforms present research gaps that leave their usage and information dissemination warranting further evaluation. Future studies should also include social platforms targeting non-English users to provide a wider global perspective.

(J Med Internet Res 2022;24(1):e28152) doi:10.2196/28152

KEYWORDS

COVID-19; Twitter; health; social media; bibliometric; dissemination; knowledge exchange

Introduction

Public health information has been traditionally distributed to the public with the use of printed media, television, or radio. With the rise of participatory web and social media [1] and particularly in the face of recent pandemics, such as the H1N1 influenza pandemic in 2009 and the COVID-19 pandemic [2], the internet plays a major role in information sharing. The general public no longer acts as a passive consumer but plays a critical role in the generation, filtering, and amplification of public health information [1]. Health care–related scientific discoveries are now often condensed into news pieces written in layman's terms and disseminated to broad and nonexpert audiences via social media, which contributes to not only better visibility of important information, but also better communication between health care professionals and the community [3]. Another major benefit of social media for health care is the potential for patient empowerment by providing a platform where patients can get information about their medical condition, communicate with health care professionals, share their experiences, and support other individuals affected by the same condition [4].

While providing numerous empowerment opportunities, there lies a great potential for miscommunication and misinformation [5] within the social media–based setting of health-related information distribution. While social media has increased and improved the dissemination of scientific results to the community, it has also increased the sensationalist language used to describe scientific findings [6,7]. Often, media articles may report research findings with misinterpretation and overstatement that can lead to confusion, misinformation, and mistrust in scientific reporting [6]. Moreover, social media empowers pseudoexperts and nonexpert influencers in sharing opinions and false information in the area of health care [8]. Very often, important societal figures, such as celebrities, politicians, and activists, without any expert knowledge of a certain topic, but with a large influence, can take part in spreading health-related misinformation [8]. The need for social media to moderate the information shared and increase expert consultation is increasingly evident and could be one way to reduce the spread of misinformation [9].

One of the most polarizing topics in recent years has been vaccination, following a scientific article from 1998 by Wakefield et al, which proposed a causative link between the measles, mumps, and rubella (MMR) vaccine and autism [3,10]. The study by Wakefield et al was later found to be flawed and fraudulent and was retracted [11-14]. Even though the findings in the study by Wakefield et al have since been disproved as numerous subsequent studies found no link between vaccines and autism, the study caused great damage to vaccine programs worldwide, with a considerable increase in the number of people rejecting vaccination in the past decades [9]. Another prominent illustrative example is the case of measles reappearance in the United States [15]. In the United States, there was an immense surge in antivaccine Tweets around 2015 to 2016, closely following the 2014 to 2015 measles outbreak and the release of Wakefield's antivaccine movie *Vaxxed* in 2016 [16]. This could be linked to the finding that antivaccine posts on Facebook were often shared and liked more often than provaccine posts [17]. Similarly, individuals exposed to negative opinions on vaccination are more likely to further share the opinions compared to individuals exposed to positive or neutral opinions [18]. This is potentiated by the so called "echo chamber effect," where many social media users are exposed to curated content that is likely to align to their existing beliefs and exacerbates the strength of the misinformation they receive [19,20].

As medical misinformation is an increasingly relevant topic to study, the amount of available literature is growing. On this background, the aim of this study was to perform a bibliometric analysis of the current literature to discover the prevalent trends and topics related to medical misinformation on social media. Conventional systematic reviews have been published on this topic to identify original articles and summarize their methodological approaches and themes [7,21]. A bibliometric study could complement their findings, for instance, by identifying the most productive authors and institutions, evaluating the geographical distribution of the publications, revealing recurring journals disseminating such research findings, unveiling the most common keywords or concepts reported, and evaluating if the publications were well cited and disseminated in journals with high impact factors. These data can serve as starting points to guide fellow researchers to pinpoint papers relevant to their studies, contact potential collaborators to conduct joint research, and find suitable journals to submit their work. At the same time, these data can help researchers find missing gaps in the literature and missing parties contributing to the field, so that the missing pieces can be filled. Since the most common social media platforms have originated

in the United States, it was hypothesized that the United States would have the highest contribution in this area of academic research. This would be an important research question as publication bias toward the United States might shadow or fail to capture the wider spectrum of global developments and experiences regarding medical misinformation on social media.

Methods

Data Source and Search Strategy

A bibliometric analysis is a study that applies mathematical and statistical methods to books and other media of communication, such as academic publications [22]. Similar to a previous bibliometric study [23], this work was performed in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) statement [24]. The Web of Science (WoS) Core Collection database was accessed on January 13, 2021, via the following search string: ALL=(misinformati* OR "wrong informati*" OR disinformati* OR "misleading informati*" OR "fake news*") AND ALL=(medic* OR illness* OR disease* OR health* OR pharma* OR drug* OR therap*) AND ALL=("social media*" OR Facebook* OR Twitter* OR Instagram* OR YouTube* OR Weibo* OR Whatsapp* OR Reddit* OR TikTok* OR WeChat*). The PubMed database was similarly searched for papers mentioning these terms in their titles and abstracts. The search terms about misinformation and its common synonyms were referred from 2 recent systematic reviews [7,25]. No additional filter was placed to restrict the search results, and the indicated search yielded 529 papers in WoS and 285 papers in PubMed. After merging the lists from both databases and removing duplicates, 529 papers remained. Since this was a total-scale analysis of the concerned literature [26], all resultant papers were included without exclusion (Multimedia Appendix 1).

The "Analyze" function of WoS was used to provide initial descriptive statistics regarding the bibliographic data. The numbers of social media platform–specific papers were counted. The approach applied to Facebook is presented here as an example for the used evaluation strategy. In particular, we additionally searched with the following search string: ALL= Facebook* NOT (Twitter* OR Instagram* OR YouTube* OR Weibo* OR Whatsapp* OR Reddit* OR TikTok* OR WeChat*). When the original search string and this new search string were combined with the Boolean operator "AND," the resulting papers mentioned Facebook but not the other referenced social media.

Outcome Measures

We evaluated the publication and citation counts of contributors in terms of author, institution, country, and journal. We also computed the publication and citation counts of terms and keywords, and identified the top 10 most cited papers. The semantic content of the identified publications was analyzed in the following ways. Citations per paper (CPPs) were computed for terms occurring in the titles, abstracts, and keywords of the identified papers, and n-gram analysis was conducted to identify the most recurring metatext. These analyses aimed to answer the queries listed at the end of the Introduction. Further details are described below.

Data Extraction and Main Analysis

The 529 identified papers were exported in full record with cited references to VOSviewer [27,28] for subsequent bibliometric analyses and visualizations. To visualize the results, a term map was created via VOSviewer to display publication and citation data for terms that appeared in the titles and abstracts of the analyzed papers. We decided to visualize terms that appeared in over 1% of the papers (ie, at least six papers) for improved clarity of the generated image, to avoid a heavily crowded figure [26]. A keyword map was similarly produced with the same frequency threshold, displaying author keywords and keywords added by WoS (KeyWords Plus) altogether. VOSviewer performs text mining by part-of-speech tagging with the aid of Apache OpenNLP and a linguistic filter, and converts plural noun phrases into singular form [29]. Meanwhile, it constructs a map in the following 3 steps based on a co-occurrence matrix: (1) calculation of a similarity index based on association strength (also known as proximity index and probabilistic affinity index), (2) application of the VOS mapping technique to the matrix, and (3) solution transformation to produce consistent results [27]. Besides visualizations, the resultant data from the maps were checked, and the recurring items were presented in tabular format.

In addition, keyword maps were produced for subsets of papers that were specific to Twitter, YouTube, and Facebook. For these maps, keywords with at least two appearances were included.

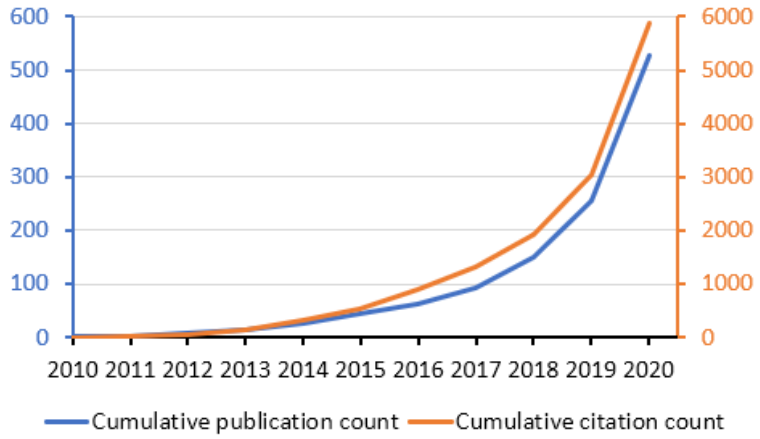
Exploratory Analysis

Finally, an exploratory n-gram analysis was conducted with the online NGram Analyzer [30] that allows n-gram metatexts to be listed. The abstracts of the publications were pasted into the program and the recurring 5-grams (a contiguous sequence of 5 words) were extracted. After manual checking, meaningful 5-grams with at least four appearances have been reported in the Results.

Results

Our search strategy identified a total of 529 scientific articles addressing medical misinformation on social media. The analysis of these papers revealed that the earliest papers on this subject could be traced back to 2010 and 2011, and the total publication and citation counts increased very rapidly, especially during the last 2 years (Figure 1). Original articles accounted for the majority of the identified publications (n=393, 74.3%), followed by editorial materials (n=50, 9.5%). The article-to-review ratio was 12.7:1 (n=393 vs 31). Proceedings accounted for another 7.2% (n=38). Over 97% of the indexed papers were written in English. The most cited paper among the 529 was also the oldest paper; it involved content analysis of over 5000 relevant Tweets during the 2009 H1N1 outbreak [1]. Within a decade, it has already accumulated 589 citations.

Figure 1. Total publication and citation counts of papers on medical and health-related misinformation on social media. Data are shown until the end of 2020.



The most productive author publishing in this subject area was Emily K Vraga from George Mason University (Virginia, USA). She started to publish on this topic in 2015 and accumulated a total of 13 papers, mostly with Leticia Bode and Melissa Tully. Leticia Bode and Joseph A Hill followed second in the list of the most productive researchers, with 9 papers each. Following them were 27 authors with 7 papers each. The top 10 most productive institutions, countries, journals, and WoS categories in which the analyzed works were published are listed in Table 1. The United States contributed to half (265/529, 50.1%) of the identified papers and was the home country of 80% of the top 10 most productive institutions. The identified papers were mostly published in journals belonging to the following categories: public environmental and occupational health, communication, health care sciences services, medical informatics, and medicine general internal.

Social media platform–specific papers accounted for 44.1% (n=233) of all 529 identified papers (Table 2). The most popularly investigated social media were Twitter, YouTube, and Facebook. They also had higher CPPs than other social media.

Table 1. Top 10 most productive institutions, countries, journals, and Web of Science categories publishing papers on medical and health-related misinformation on social media.

Variable	Publication count (N=529), n (%)	Citations per paper
Institution		
Harvard University	25 (4.7)	13.2
University of Texas System	20 (3.8)	3.4
University of North Carolina	14 (2.6)	13.8
University of Pennsylvania	14 (2.6)	11.9
University of London	13 (2.5)	30.5
Johns Hopkins University	12 (2.3)	6.0
University of California System	11 (2.1)	1.7
University of Minnesota System	11 (2.1)	2.8
Pennsylvania Commonwealth System of Higher Education	10 (1.9)	3.6
University System of Sydney	10 (1.9)	32.1
Country		
United States	265 (50.1)	12.2
United Kingdom	53 (9.3)	20.0
Italy	35 (6.6)	9.2
Canada	33 (6.2)	34.0
Spain	30 (5.7)	7.2
Australia	27 (5.1)	19.0
China	27 (5.1)	13.7
Turkey	17 (3.2)	5.1
Germany	15 (2.8)	27.9
India	14 (2.6)	2.8
Switzerland	14 (2.6)	14.9
Journal (2019 impact factor)		
Journal of Medical Internet Research (5.034)	32 (6.0)	14.1
American Journal of Public Health (6.464)	14 (2.6)	3.1
Health Communication (1.965)	13 (2.5)	9.2
Vaccine (3.143)	13 (2.5)	28.1
International Journal of Environmental Research and Public Health (2.468)	11 (2.1)	8.6
PLOS One (2.740)	11 (2.1)	60.1
Annals of Behavioral Medicine (4.475)	8[a] (1.5)	0
Professional de la Informacion (N/A[b])	8 (1.5)	8.6
Cureus (N/A)	6 (1.1)	26.7
Journal of Health Communication (1.596)	6 (1.1)	2.3
Web of Science category		
Public environmental and occupational health	95 (18.0)	12.6
Communication	71 (13.4)	7.3
Health care sciences services	50 (9.5)	17.5
Medical informatics	48 (9.1)	17.4
Medicine general internal	38 (7.2)	17.2

Variable	Publication count (N=529), n (%)	Citations per paper
Computer science information systems	33 (6.2)	4.1
Information science library science	32 (6.0)	5.5
Health policy services	22 (4.2)	13.6
Computer science theory methods	21 (4.0)	5.4
Immunology	21 (4.0)	23.6

[a]All 8 publications in Annals of Behavioral Medicine were meeting abstracts and received no citation.

[b]N/A: not applicable.

Table 2. Count of platform-specific papers on medical and health-related misinformation on social media.

Social media	Publication count, n	Citations per paper
Twitter	90	17.0
YouTube	67	13.7
Facebook	57	15.3
WhatsApp	6	4.0
Instagram	6	8.7
Weibo	4	7.5
Reddit	2	2.5
WeChat	1	3.0
TikTok	0	N/A[a]

[a]N/A: not applicable.

Figure 2 shows the terms extracted from the titles and abstracts of all 529 identified papers. COVID-19 (“covid” at the lower half, n=109, CPP=7.1) and vaccine (upper half, n=62, CPP=15.7) were 2 major health issues identified. Mentioned COVID-19 derivatives included SARS-CoV-2 (“sars cov,” n=9, CPP=11.0), coronavirus (n=22, CPP=15.3), coronavirus disease (n=25, CPP=12.6), and coronavirus pandemic (n=6, CPP=2.3). Mentioned vaccine derivatives included vaccination (n=53, CPP=21.6), vaccination rate (n=7, CPP=5.7), vaccine hesitancy (n=14, CPP=13.8), vaccine misinformation (n=13, CPP=6.9), vaccine preventable disease (n=6, CPP=29.0), and vaccine safety (n=8, CPP=7.8). The top 20 terms with the highest CPPs are listed in Table 3. Notable terms hinting on important issues discussed in the analyzed literature set were public perception, public concern, health authority, peer (related to peer-to-peer support), and policy maker (Table 3).

Figure 2. Term map showing words/phrases extracted from the titles and abstracts of the 529 papers. Circle size is related to the number of papers mentioning the word/phrase. Circle color is related to the citations per paper. The proximity between circles is related to how frequently the terms are co-mentioned in the same papers.

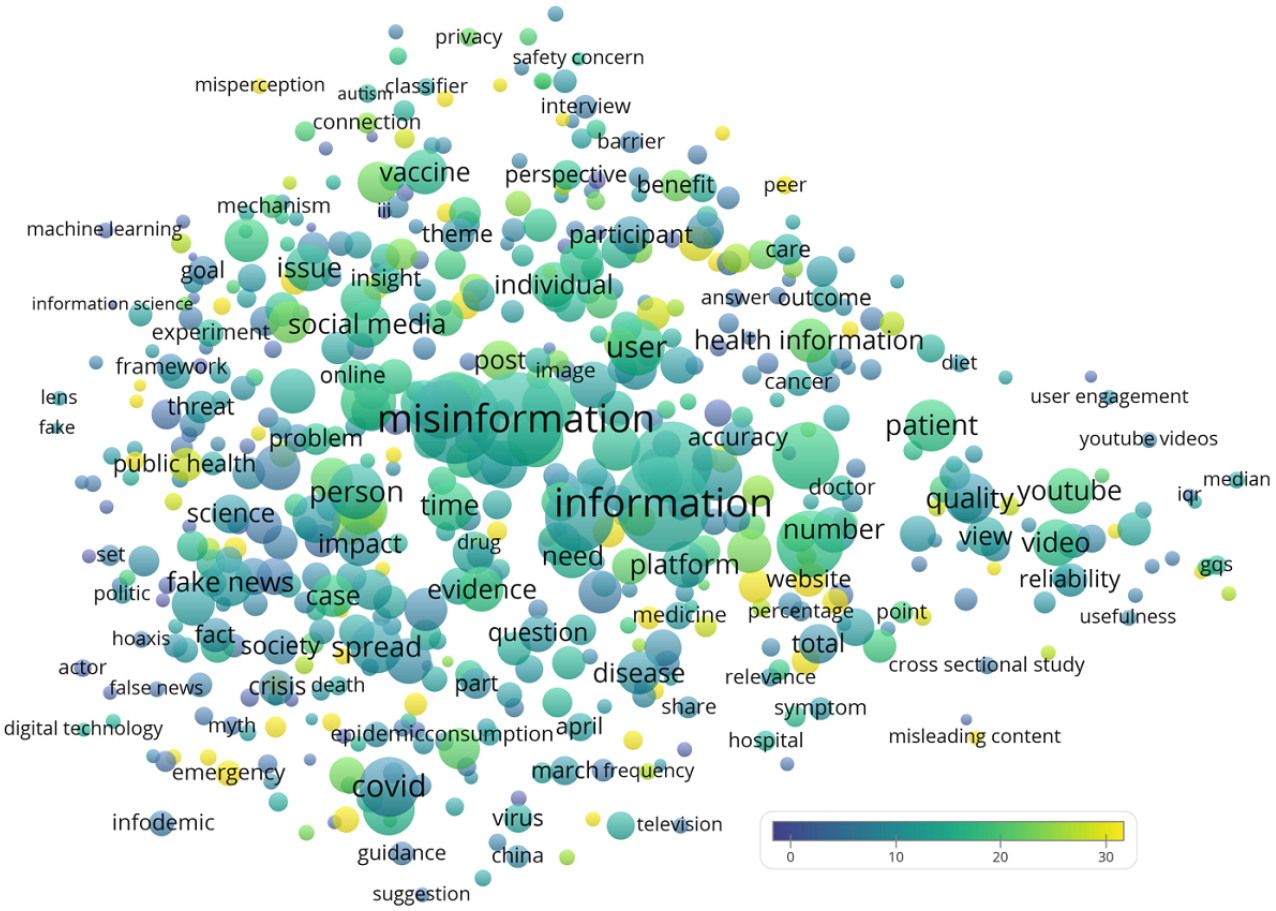


Table 3. Top 20 terms with the highest citations per paper.

Term[a]	Publication count (N=529), n (%)	Citations per paper
Real time	5 (0.9)	160.2
Public perception	7 (1.3)	86.4
Credible source	7 (1.3)	84.9
Public concern	11 (2.1)	75.5
Health authority	14 (2.6)	56.5
Story	24 (4.5)	54.5
Peer	11 (2.1)	50.4
Adoption	16 (3.0)	49.3
Relevant video	7 (1.3)	48.1
Term	34 (6.4)	43.3
Sentiment	19 (3.6)	41.2
Illness	6 (1.1)	41.0
Zika virus	6 (1.1)	40.3
Emergency	21 (4.0)	38.7
Policy maker	9 (1.7)	38.6
Viewer	7 (1.3)	37.1
Misperception	8 (1.5)	36.5
Information source	12 (2.3)	36.0
Feeling	6 (1.1)	35.0
Potential risk	7 (1.3)	35.0

[a]Only terms that appeared in at least 1% of papers were considered.

A keyword map is shown in Figure 3. The keyword map showed that several diseases were recurring themes of investigation, such as measles (n=9, CPP=7.7), Ebola (n=22, CPP=11.4), COVID-19 (n=87, CPP=6.4), and cardiovascular diseases (n=9, CPP=1.7). Table 4 presents the top 20 keywords with the highest CPPs, and reveals that risk and safety were among the concepts with the highest CPPs.

Figure 3. Keyword map of the 529 papers. Circle size is related to the number of papers including the word/phrase as a keyword. Circle color is related to the citations per paper. The proximity between circles is related to how frequently the terms are co-mentioned in the same papers.

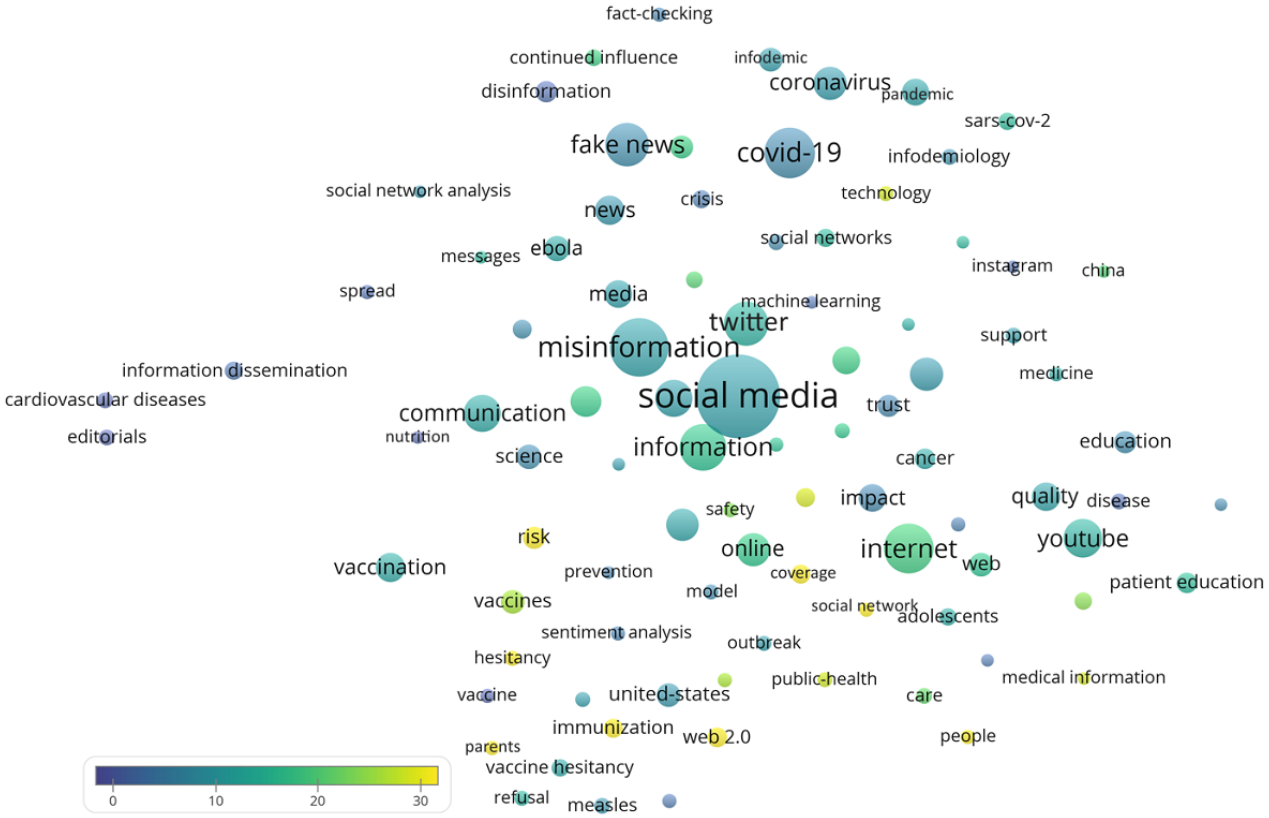


Table 4. Top 20 keywords with the highest citations per paper.

Keyword[a]	Publication count (N=529), n (%)	Citations per paper
Risk	17 (3.2)	51.6
Social network	7 (1.3)	50.7
Parents	7 (1.3)	41.1
Hesitancy	8 (1.5)	39.5
Coverage	13 (2.5)	37.9
Immunization	13 (2.5)	34.5
People	7 (1.3)	31.0
Web 2.0	14 (2.6)	30.4
Knowledge	13 (2.5)	27.5
Medical information	6 (1.1)	27.5
Technology	8 (1.5)	27.3
Public-health	8 (1.5)	27.1
Attitudes	7 (1.3)	24.9
Vaccines	19 (3.6)	24.7
Videos	11 (2.1)	23.1
Safety	7 (1.3)	22.7
Care	9 (1.7)	20.9
Risk communication	10 (1.9)	20.0
China	6 (1.1)	19.5
Internet	86 (16.3)	18.6

[a]Only keywords that appeared in at least 1% of papers were considered.

Keyword maps generated for the publication set that investigated Twitter, YouTube, and Facebook are shown in Figure 4A-C. The keyword maps reveal that Twitter research focused on anxiety related to online searching for disease and medical information (cyberchondria and hypochondriasis, cyan), vaccination and Zika virus (red), COVID-19 (blue), Ebola (yellow), cancer (green), and data analysis involving predictive modeling (purple). YouTube research focused on smoking and tobacco (purple), alternative medicine for various diseases, such as rheumatoid arthritis and prostate cancer (green), breast cancer (yellow), COVID-19 (red), and Ebola (blue). Finally, Facebook research focused on online health communities (yellow), vaccine hesitancy related to autism (blue), credibility of health information related to immunization and nutrition (red), cancer (purple), and COVID-19 (green).

The top 10 most cited papers are listed in Table 5. Peer-to-peer support and spread of misinformation were mentioned, and all Twitter, YouTube, and Facebook data were investigated. The themes of these top 10 papers were consistent with the list of highly cited terms listed in Table 3, covering topics such as peer-to-peer support, online videos, and public perception.

The exploratory n-gram analysis resulted in several meaningful 5-gram metatexts with at least four appearances as follows: 6 appearances, “as a source of information” and “the spread of fake news;” 5 appearances, “rumors stigma and conspiracy theories,” “the quality of health information,” and “content reliability and quality of;” 4 appearances, “health anxiety and health literacy,” “the relationship between message factors,” “intentions to trust and share,” “#PedsICU and coronavirus disease 2019,” “in low- and middle-income countries,” “actions for a framework for,” “interacted with perceived message importance,” and “verify and share the message.”

Figure 4. Keyword maps of the papers investigating (A) Twitter, (B) YouTube, and (C) Facebook. Circle size is related to the number of papers mentioning the respective word/phrase as a keyword. Circle color is related to the clustering of the words by the default setting of VOSviewer. The proximity between circles is related to how frequently the terms are co-mentioned in the same papers.

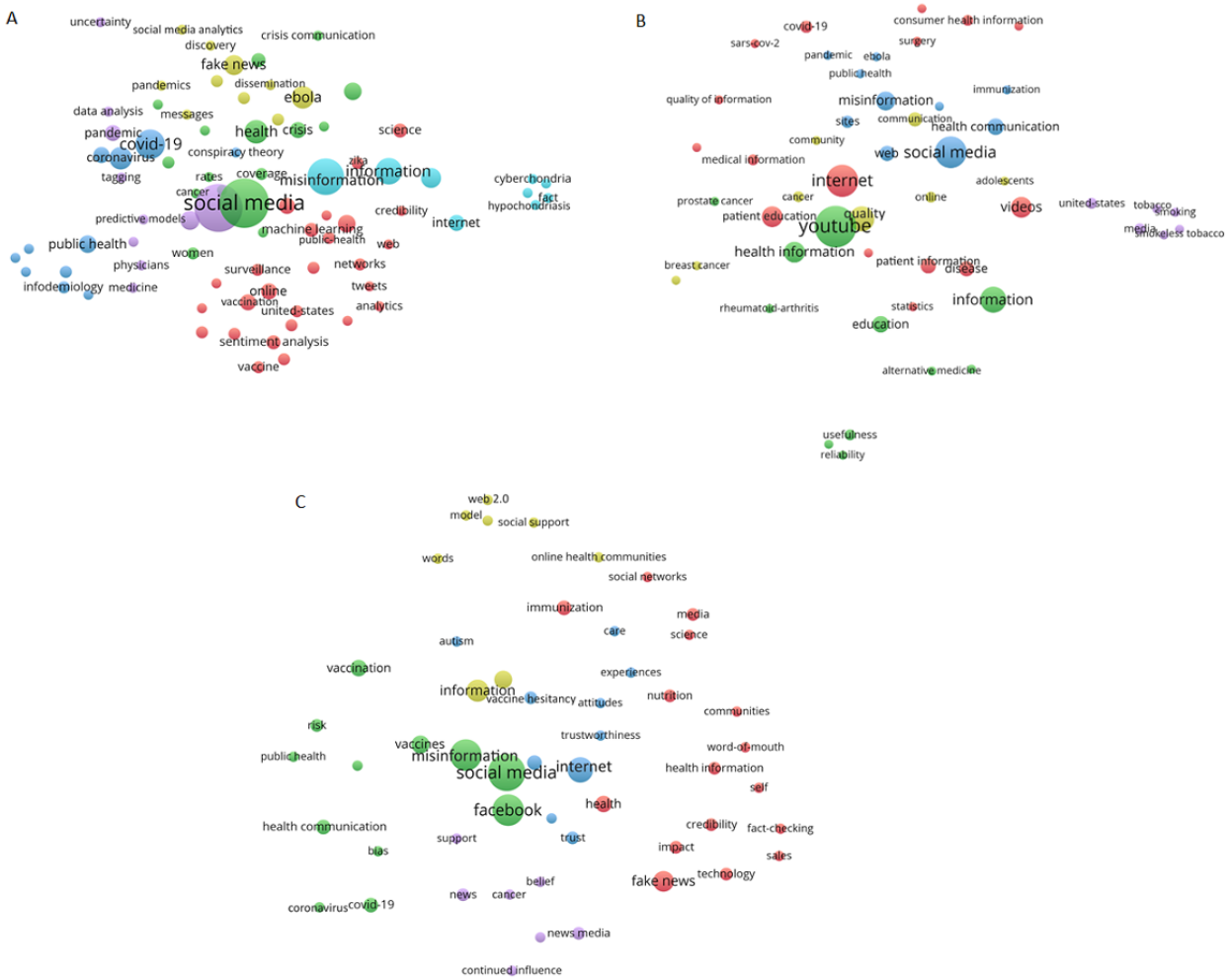


Table 5. Top 10 most cited papers on medical and health-related misinformation on social media.

Authors, year	Citations, n
Chew et al, 2010 [1]	613
Yaqub et al, 2014 [31]	256
Naslund et al, 2016 [32]	212
Madathil et al, 2015 [33]	195
Kamel Boulos et al, 2011 [34]	186
Betsch et al, 2012 [35]	168
Syed-Abdul et al, 2013 [36]	147
Depoux et al, 2020 [37]	136
Singh et al, 2012 [38]	123
Bode et al, 2015 [39]	121

Discussion

General Discussion

Using bibliometric analysis, this study identified and quantitatively analyzed 529 papers on medical and health-related misinformation on social media, revealing the most popularly investigated social media platforms, prevailing research themes, most utilized journals, and most productive countries, institutions, and authors.

Findings Concerning the Western World and Its Prevalent Social Media Platforms

This bibliometric analysis on 529 scientific articles concerning medical and health-related misinformation on social media revealed that the most heavily investigated platforms were Twitter, YouTube, and Facebook. This could be related to the finding that most of the top 10 productive countries were from North America and Europe where these social media platforms were dominant. The results also confirmed the hypothesis that the United States had the largest contribution in social media research. The total publication and citation counts increased very rapidly especially during the last 2 years, consistent with the trends identified by previous systematic reviews on this topic [7,21]. On the other hand, this study found that original articles accounted for 74.3% of the analyzed literature set. This implied that one-fourth of the literature was not covered by the abovementioned systematic reviews, which might partly explain some differences in the results. For instance, this study found that Twitter was the most recurring social medium in the literature instead of YouTube, as reported by Wang et al [7]. The strength of the review by Suarez-Lledo and Alvarez-Galvez [21] was that it analyzed the prevalence of health misinformation posts (0%-98%) reported in the original articles. Meanwhile, Wang et al [7] categorized them into predefined theoretical frameworks with the most prevalent ones being public health, health policy, and epidemiology (n=14); health informatics (n=8); communications studies (n=5); vaccines (n=4); cyberpsychology (n=3); and system sciences (n=3). Here, it was found that publications in immunology were on average more frequently cited than communication and computer science, whereas health care sciences and medical informatics papers were in-between. This implies that the more published disciplines do not necessarily warrant more citations. This finding has 2 implications. First, quantity may not necessarily mean quality. Second, field differences in citation rates found in general [40] remained present even when the literature set was confined to a single focus on misinformation. Similar to the current findings, Wang et al [7] also found that the most popular topics were vaccination, Ebola, and Zika virus, with other less popular focus topics being nutrition, cancer, and smoking. In contrast, Wang et al [7] identified fluoridation of water as one of the recurring topics, whereas in this study, COVID-19 emerged as a strong research focus. Moreover, the performed keyword analysis in this work revealed that the research on different social media platforms, such as Twitter, YouTube, and Facebook, focused on different topics. While, at present, we do not have an explanation for this interesting observation, we believe that the reasoning for different topic studies on different social media could be a relevant direction for future research. Such studies may elucidate whether this is due to different prevalences of specific content across the platforms or due to preferential academic interest from research teams with particular interest in specific social media platforms or topics.

Findings Concerning China and Its Prevalent Social Media Platforms

Though some social media platforms were not available in China, China still made it into the top 10 list of the most productive countries (Table 1). With a large population of internet users in China, it could be reasonably expected that Weibo and WeChat, which are popular social media platforms in China, would become more investigated in future studies. One potential barrier to non-Chinese researchers would be content translation, as the majority of their content is written in Chinese. In addition, the fast-growing short video platform TikTok (and its Chinese version Douyin) might also exert significant influence on the health information seeking behavior of internet users in the future. However, TikTok videos might be hard to archive, and video analysis tools might not be as well developed as text mining tools, which might hinder analysis by public health researchers. The same applies to the visual contents posted on Instagram. Current findings seem to suggest that sufficient research on misinformation disseminating through these platforms is missing from the current literature and should be addressed in future research. Readers should be aware that the publication bias toward Europe and North America, especially the United States, indicates that the current body of knowledge might not reflect the wider spectrum of misinformation on global health issues, especially in other parts of the world with large online communities, such as Asia and South America.

The most productive author was found to be Emily K Vraga who is based in the United States. Her studies focused on how to correct health misinformation (in other words, overturn subjects' misperceptions) dispersed on social media, particularly Facebook and Twitter [39,41-43]. Though China was among the top 10 most productive countries, we found that only 2 of the top 50 most productive authors were based in China. They were King-Wa Fu (n=4) and Chung-Hong Chan (n=3) from the University of Hong Kong, and they focused solely on the Ebola virus [44-47]. This implied that, to grasp the research foci from China, readers need to refer to diverse works from multiple authors instead of that from a few prominent authors. With the continued growth of netizens in China, we anticipate that more productive authors might be based in China in the future.

Elaboration on the Recurring Themes of the Literature

A very important role of social media is to provide peer-to-peer support, as investigated by some publications identified in this study (see Tables 3 and 5), for example, by forming online health communities and support groups, and ensuring stakeholder access to the latest and most relevant scientific information and health interventions [32]. For instance, users could post supportive comments and advice to YouTube videos uploaded by individuals with mental health issues [48]. On the other hand, misinformation spread via social media (especially related to Twitter, see Figure 4A) might lead to cyberchondria (cyberspace-related hypochondria; the unfounded concern escalation about common symptomology based on information found on the internet), with a study revealing that unverified information might be more easily shared by internet users who trust online information and perceived an information overload, and that women were more likely to experience cyberchondria [49]. Cyberchondria could be an important health concern, as a meta-analysis established the correlation between health anxiety and both online information seeking and cyberchondria [50], and another work revealed that it had 5 facets, including

reassurance seeking and mistrust of medical professionals [51]. Being flooded by online misinformation would not alleviate the situation but may worsen it. A recent study found that government and professional videos containing solely factual information only accounted for 11% of all YouTube videos on COVID-19 and 10% of views, whereas 28% of the analyzed videos contained misinformation and had up to 62 million views [52]. In this context, the adequacy of funding and resources allocated by governmental bodies to online health literacy campaigns needs to be questioned.

Meanwhile, from the perspective of policy makers, the large amount of information from social media can be monitored and used for the achievement of efficient outcomes. As seen from the results, "health policy services" was among the most recurring journal categories for the analyzed literature set (Table 1) and "policy maker" was one of the recurring terms with the highest CPP (Table 3). For instance, by analyzing tweets related to the COVID-19 pandemic, researchers could identify the top 4 concerns of Twitter users, namely the origin, sources, impact on various levels, and ways to reduce the infection risk [53]. While using similar approaches, keeping these concerns anonymized at an individual level and ensuring that social or ethnic groups expressing specific concerns do not become targets of discrimination are crucial. Authorities could therefore focus on these concerns as they derive measures and disseminate information to the public to contain the pandemic and reduce fears within the community. In this regard, authorities could collaborate with scientific associations and provide incentives to civil society to address ignorance or misinformation on the detected concerns. Future research could compare relevant social media content following interventions, to define the optimal strategies of tackling misinformation on social media.

As mentioned in the Introduction, the fraudulent study linking the MMR vaccine to autism still has a lingering influence on social media, as it is still posted on the Facebook platform by antivaccine advocates, despite its retraction due to fraudulency [54]. Moreover, it was found that the content posted by Facebook users regarding vaccination has been increasingly polarized [20]. One study suggested that the use of angry language could promote the viral spread of the messages, including the misinformation of vaccines causing autism [55], though it was not investigating contents exclusive to vaccines and only binarized words into positive and negative emotions. Summarizing the results from n-gram analysis, netizens might need to be aware of fake news, rumors, stigma, and conspiracy theories circulating on the internet. Content reliability and quality should be assessed, and information should be verified before sharing. One way to cohort authoritative or accurate health care information shared by experts on social media (eg, Twitter) is by the use of a hashtag, so that others can search easily. One example was #PedsICU that promoted pediatric critical care content, as found by n-gram analysis. By sensible collaboration, there may be a chance to mitigate misinformation.

Limitations

Any publications in journals not indexed by WoS were missed in the current analysis. For example, there is a relevant paper investigating misinformation of COVID-19 on TikTok, which is not indexed by WoS [56]. Besides, WoS mainly indexes papers written in English. There may be papers investigating Weibo and WeChat written in Chinese or published in Chinese journals that are not yet indexed by WoS. Preprints are also not indexed in WoS, which could be an important source of preliminary information, but the reliability of such information is debatable due to the lack of peer-review assessment. Moreover, a bibliometric study cannot assess the scientific quality of the content, such as risk of bias, effect size or statistical significance of the results, and whether the conclusions are justified by the respective data reported. The accuracy of data tagging by the literature database could also pose a limitation. For instance, KeyWords Plus are keywords tagged to a paper by an algorithm used by WoS based on the terms from the titles of the cited references [57], and are more broadly descriptive and therefore applicable to analyzing the structure of scientific fields [58]. However, it was unclear how accurate they were compared to other tags such as the National Center for Biotechnology Information's Medical Subject Headings ("MeSH terms"). Future studies should also incorporate "conspiracy theory" and related terms into their search protocols for more comprehensive results.

Future Research

For potential future research, artificial intelligence (AI) applications for social media content analysis would be an especially promising avenue. With increasing content and misinformation circulating on social media, it becomes practically impossible to manually determine and classify misinformation. AI or machine learning might be employed for such content analysis, which has the potential to achieve high accuracy [59]. Yet, AI could also be exploited to generate and disseminate misinformation to targeted audiences [60]. AI research in health care was most frequently published in journals in computer science and engineering, as reported by Guo et al [61], and indeed, among their identified top 10 most productive journals, only PLOS One was on the list in our study (Table 1). Along this line, with the further development of AI applications for social media content analysis, it might also be of interest to promote the dissemination of such research in mainstream public health journals, in order to reach a broader relevant audience.

Conclusions

Based on an analysis of 529 papers on medical and health-related misinformation on social media, we found that the United States contributed to half of the papers, with 80% of the top 10 most productive institutions being based in this country. The papers were mostly published in journals belonging to the categories public environmental and occupational health, communication, health care sciences services, medical informatics, and medicine general internal. However, they were generally less cited than papers published in immunology, suggesting that more publications did not warrant more citations. Social media platform–specific papers accounted for 44% of all papers. The most popularly investigated social media platforms were Twitter, YouTube, and Facebook. They also had higher CPPs than other social media. Investigations on these platforms had different foci. Twitter-based research investigated cyberchondria and hypochondriasis, YouTube-based research

investigated tobacco smoking, and Facebook-based research investigated vaccine hesitancy related to autism. COVID-19 was a common topic investigated across all platforms. An important implication of these findings is that often knowledge on specific themes related to medical misinformation relies on the predominant study of a single social media platform or limited number of platforms, and broader cross-platform studies could be a promising direction for future research. Future studies should also include social platforms aimed at non-English users to provide a wider perspective on global health misinformation.

Conflicts of Interest

None declared.

Multimedia Appendix 1
PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analysis) flow diagram of the literature search.
[PNG File , 49 KB - jmir_v24i1e28152_app1.png]

References

1. Chew C, Eysenbach G. Pandemics in the age of Twitter: content analysis of Tweets during the 2009 H1N1 outbreak. PLoS One 2010 Nov 29;5(11):e14118 [FREE Full text] [doi: 10.1371/journal.pone.0014118] [Medline: 21124761]
2. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
3. Jang SM, Mckeever BW, Mckeever R, Kim JK. From Social Media to Mainstream News: The Information Flow of the Vaccine-Autism Controversy in the US, Canada, and the UK. Health Commun 2019 Jan 13;34(1):110-117. [doi: 10.1080/10410236.2017.1384433] [Medline: 29028371]
4. Househ M, Borycki E, Kushniruk A. Empowering patients through social media: the benefits and challenges. Health Informatics J 2014 Mar 18;20(1):50-58 [FREE Full text] [doi: 10.1177/1460458213476969] [Medline: 24550564]
5. Hagg E, Dahinten VS, Currie LM. The emerging use of social media for health-related purposes in low and middle-income countries: A scoping review. Int J Med Inform 2018 Jul;115:92-105 [FREE Full text] [doi: 10.1016/j.ijmedinf.2018.04.010] [Medline: 29779724]
6. Haber N, Smith ER, Moscoe E, Andrews K, Audy R, Bell W, CLAIMS research team. Causal language and strength of inference in academic and media articles shared in social media (CLAIMS): A systematic review. PLoS One 2018 May 30;13(5):e0196346 [FREE Full text] [doi: 10.1371/journal.pone.0196346] [Medline: 29847549]
7. Wang Y, McKee M, Torbica A, Stuckler D. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. Soc Sci Med 2019 Nov;240:112552 [FREE Full text] [doi: 10.1016/j.socscimed.2019.112552] [Medline: 31561111]
8. Racovita M. Lost in translation. Scientists need to adapt to a postmodern world; constructivism can offer a way. EMBO Rep 2013 Aug 16;14(8):675-678 [FREE Full text] [doi: 10.1038/embor.2013.90] [Medline: 23856722]
9. Hill J, Agewall S, Baranchuk A, Booz GW, Borer JS, Camici PG, et al. Medical Misinformation: Vet the Message!. J Am Heart Assoc 2019 Feb 05;8(3):e011838 [FREE Full text] [doi: 10.1161/JAHA.118.011838] [Medline: 30686129]
10. Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson DM, Malik M, et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. Lancet 1998 Feb 28;351(9103):637-641. [doi: 10.1016/s0140-6736(97)11096-0] [Medline: 9500320]
11. No Authors. A timeline of the Wakefield retraction. Nat Med 2010 Mar;16(3):248-248. [doi: 10.1038/nm0310-248b]
12. Dyer C. Lancet retracts Wakefield's MMR paper. BMJ 2010 Feb 02;340(feb02 4):c696-c696. [doi: 10.1136/bmj.c696] [Medline: 20124366]
13. Godlee F, Smith J, Marcovitch H. Wakefield's article linking MMR vaccine and autism was fraudulent. BMJ 2011 Jan 05;342(jan05 1):c7452-c7452. [doi: 10.1136/bmj.c7452] [Medline: 21209060]
14. Omer SB. The discredited doctor hailed by the anti-vaccine movement. Nature 2020 Oct 27;586(7831):668-669. [doi: 10.1038/d41586-020-02989-9]
15. Hotez PJ, Nuzhath T, Colwell B. Combating vaccine hesitancy and other 21st century social determinants in the global fight against measles. Curr Opin Virol 2020 Apr;41:1-7. [doi: 10.1016/j.coviro.2020.01.001] [Medline: 32113136]
16. Gunaratne K, Coomes EA, Haghbayan H. Temporal trends in anti-vaccine discourse on Twitter. Vaccine 2019 Aug 14;37(35):4867-4871. [doi: 10.1016/j.vaccine.2019.06.086] [Medline: 31300292]
17. Gandhi CK, Patel J, Zhan X. Trend of influenza vaccine Facebook posts in last 4 years: a content analysis. Am J Infect Control 2020 Apr;48(4):361-367. [doi: 10.1016/j.ajic.2020.01.010] [Medline: 32098695]
18. Dunn AG, Leask J, Zhou X, Mandl KD, Coiera E. Associations Between Exposure to and Expression of Negative Opinions About Human Papillomavirus Vaccines on Social Media: An Observational Study. J Med Internet Res 2015 Jun 10;17(6):e144 [FREE Full text] [doi: 10.2196/jmir.4343] [Medline: 26063290]

19. Chou WS, Oh A, Klein WMP. Addressing Health-Related Misinformation on Social Media. JAMA 2018 Dec 18;320(23):2417-2418. [doi: 10.1001/jama.2018.16865] [Medline: 30428002]
20. Schmidt AL, Zollo F, Scala A, Betsch C, Quattrociocchi W. Polarization of the vaccination debate on Facebook. Vaccine 2018 Jun 14;36(25):3606-3612. [doi: 10.1016/j.vaccine.2018.05.040] [Medline: 29773322]
21. Suarez-Lledo V, Alvarez-Galvez J. Prevalence of Health Misinformation on Social Media: Systematic Review. J Med Internet Res 2021 Jan 20;23(1):e17187 [FREE Full text] [doi: 10.2196/17187] [Medline: 33470931]
22. Pritchard A. Statistical bibliography or bibliometrics. Journal of Documentation 1969;25(4):348-349.
23. Huang Z, Chen H, Liu Z. The 100 top-cited systematic reviews/meta-analyses in central venous catheter research: A PRISMA-compliant systematic literature review and bibliometric analysis. Intensive Crit Care Nurs 2020 Apr;57:102803. [doi: 10.1016/j.iccn.2020.102803] [Medline: 32035805]
24. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med 2009 Jul 21;6(7):e1000097 [FREE Full text] [doi: 10.1371/journal.pmed.1000097] [Medline: 19621072]
25. Celliers M, Hattingh M. A Systematic Review on Fake News Themes Reported in Literature. In: Hattingh M, Matthee M, Smuts H, Pappas I, Dwivedi YK, Mäntymäki M, editors. Responsible Design, Implementation and Use of Information and Communication Technology. I3E 2020. Lecture Notes in Computer Science, vol 12067. Cham: Springer; 2020:223-234.
26. Atanasov AG, Yeung AWK, Klager E, Eibensteiner F, Schaden E, Kletecka-Pulker M, et al. First, Do No Harm (Gone Wrong): Total-Scale Analysis of Medical Errors Scientific Literature. Front Public Health 2020 Oct 16;8:558913 [FREE Full text] [doi: 10.3389/fpubh.2020.558913] [Medline: 33178657]
27. van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. Scientometrics 2010 Aug 31;84(2):523-538 [FREE Full text] [doi: 10.1007/s11192-009-0146-3] [Medline: 20585380]
28. Waltman L, van Eck NJ, Noyons EC. A unified approach to mapping and clustering of bibliometric networks. Journal of Informetrics 2010 Oct;4(4):629-635. [doi: 10.1016/j.joi.2010.07.002]
29. Van Eck N, Waltman L. Text mining and visualization using VOSviewer. ISSI Newsletter 2011;7:50-54.
30. Online NGram Analyzer. Guide to Data Mining. URL: http://guidetodatamining.com/ngramAnalyzer/ [accessed 2022-01-11]
31. Yaqub O, Castle-Clarke S, Sevdalis N, Chataway J. Attitudes to vaccination: a critical review. Soc Sci Med 2014 Jul;112:1-11 [FREE Full text] [doi: 10.1016/j.socscimed.2014.04.018] [Medline: 24788111]
32. Naslund JA, Aschbrenner KA, Marsch LA, Bartels SJ. The future of mental health care: peer-to-peer support and social media. Epidemiol Psychiatr Sci 2016 Jan 08;25(2):113-122. [doi: 10.1017/s2045796015001067]
33. Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: A systematic review. Health Informatics J 2015 Sep;21(3):173-194 [FREE Full text] [doi: 10.1177/1460458213512220] [Medline: 24670899]
34. Kamel Boulos MN, Resch B, Crowley DN, Breslin JG, Sohn G, Burtner R, et al. Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: trends, OGC standards and application examples. Int J Health Geogr 2011 Dec 21;10:67 [FREE Full text] [doi: 10.1186/1476-072X-10-67] [Medline: 22188675]
35. Betsch C, Brewer NT, Brocard P, Davies P, Gaissmaier W, Haase N, et al. Opportunities and challenges of Web 2.0 for vaccination decisions. Vaccine 2012 May 28;30(25):3727-3733. [doi: 10.1016/j.vaccine.2012.02.025] [Medline: 22365840]
36. Syed-Abdul S, Fernandez-Luque L, Jian WS, Li YC, Crain S, Hsu MH, et al. Misleading health-related information promoted through video-based social media: anorexia on YouTube. J Med Internet Res 2013 Feb 13;15(2):e30 [FREE Full text] [doi: 10.2196/jmir.2237] [Medline: 23406655]
37. Depoux A, Martin S, Karafillakis E, Preet R, Wilder-Smith A, Larson H. The pandemic of social media panic travels faster than the COVID-19 outbreak. J Travel Med 2020 May 18;27(3):taaa031 [FREE Full text] [doi: 10.1093/jtm/taaa031] [Medline: 32125413]
38. Singh AG, Singh S, Singh PP. YouTube for information on rheumatoid arthritis--a wakeup call? J Rheumatol 2012 May;39(5):899-903. [doi: 10.3899/jrheum.111114] [Medline: 22467934]
39. Bode L, Vraga EK. In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media. J Commun 2015 Jun 23;65(4):619-638. [doi: 10.1111/jcom.12166]
40. Bornmann L, Haunschild R, Mutz R. Should citations be field-normalized in evaluative bibliometrics? An empirical analysis based on propensity score matching. Journal of Informetrics 2020 Nov;14(4):101098. [doi: 10.1016/j.joi.2020.101098]
41. Bode L, Vraga EK. See Something, Say Something: Correction of Global Health Misinformation on Social Media. Health Commun 2018 Sep 16;33(9):1131-1140. [doi: 10.1080/10410236.2017.1331312] [Medline: 28622038]
42. Vraga EK, Bode L. Using Expert Sources to Correct Health Misinformation in Social Media. Science Communication 2017 Sep 14;39(5):621-645. [doi: 10.1177/1075547017731776]
43. Vraga EK, Bode L. I do not believe you: how providing a source corrects health misperceptions across social media platforms. Information, Communication & Society 2017 Apr 19;21(10):1337-1353. [doi: 10.1080/1369118x.2017.1313883]
44. Fung IC, Duke CH, Finch KC, Snook KR, Tseng P, Hernandez AC, et al. Ebola virus disease and social media: A systematic review. Am J Infect Control 2016 Dec 01;44(12):1660-1671. [doi: 10.1016/j.ajic.2016.05.011] [Medline: 27425009]

45. Fung I, Fu K, Chan C, Chan B, Cheung C, Abraham T, et al. Social Media's Initial Reaction to Information and Misinformation on Ebola, August 2014: Facts and Rumors. Public Health Rep 2016;131(3):461-473 [FREE Full text] [doi: 10.1177/003335491613100312] [Medline: 27252566]
46. Fung IC, Fu K, Chan C, Chan BSB, Cheung C, Abraham T, et al. Social Media's Initial Reaction to Information and Misinformation on Ebola, August 2014: Facts and Rumors. Public Health Rep 2016 May 01;131(3):461-473 [FREE Full text] [doi: 10.1177/003335491613100312] [Medline: 27252566]
47. Liang H, Fung IC, Tse ZTH, Yin J, Chan C, Pechta LE, et al. How did Ebola information spread on twitter: broadcasting or viral spreading? BMC Public Health 2019 Apr 25;19(1):438 [FREE Full text] [doi: 10.1186/s12889-019-6747-8] [Medline: 31023299]
48. Naslund JA, Grande SW, Aschbrenner KA, Elwyn G. Naturally occurring peer support through social media: the experiences of individuals with severe mental illness using YouTube. PLoS One 2014 Oct 15;9(10):e110171 [FREE Full text] [doi: 10.1371/journal.pone.0110171] [Medline: 25333470]
49. Laato S, Islam AKMN, Islam MN, Whelan E. What drives unverified information sharing and cyberchondria during the COVID-19 pandemic? European Journal of Information Systems 2020 Jun 07;29(3):288-305. [doi: 10.1080/0960085x.2020.1770632]
50. McMullan RD, Berle D, Arnáez S, Starcevic V. The relationships between health anxiety, online health information seeking, and cyberchondria: Systematic review and meta-analysis. J Affect Disord 2019 Feb 15;245:270-278. [doi: 10.1016/j.jad.2018.11.037] [Medline: 30419526]
51. McElroy E, Shevlin M. The development and initial validation of the cyberchondria severity scale (CSS). J Anxiety Disord 2014 Mar;28(2):259-265. [doi: 10.1016/j.janxdis.2013.12.007] [Medline: 24508033]
52. Li HO, Bailey A, Huynh D, Chan J. YouTube as a source of information on COVID-19: a pandemic of misinformation? BMJ Glob Health 2020 May 14;5(5):e002604 [FREE Full text] [doi: 10.1136/bmjgh-2020-002604] [Medline: 32409327]
53. Abd-Alrazaq A, Alhuwail D, Househ M, Hamdi M, Shah Z. Top Concerns of Tweeters During the COVID-19 Pandemic: Infoveillance Study. J Med Internet Res 2020 Apr 21;22(4):e19016 [FREE Full text] [doi: 10.2196/19016] [Medline: 32287039]
54. Shelby A, Ernst K. Story and science: how providers and parents can utilize storytelling to combat anti-vaccine misinformation. Hum Vaccin Immunother 2013 Aug 27;9(8):1795-1801 [FREE Full text] [doi: 10.4161/hv.24828] [Medline: 23811786]
55. Bail CA. Emotional Feedback and the Viral Spread of Social Media Messages About Autism Spectrum Disorders. Am J Public Health 2016 Jul;106(7):1173-1180. [doi: 10.2105/ajph.2016.303181]
56. Basch C, Hillyer G, Jaime C. COVID-19 on TikTok: harnessing an emerging social media platform to convey important public health messages. Int J Adolesc Med Health 2020 Aug 10:2020. [doi: 10.1515/ijamh-2020-0111] [Medline: 32776899]
57. Garfield E, Sher IH. KeyWords Plus™—algorithmic derivative indexing. J. Am. Soc. Inf. Sci 1993 Jun;44(5):298-299. [doi: 10.1002/(sici)1097-4571(199306)44:5<298::aid-asi5>3.0.co;2-a]
58. Zhang J, Yu Q, Zheng F, Long C, Lu Z, Duan Z. Comparing keywords plus of WOS and author keywords: A case study of patient adherence research. J Assn Inf Sci Tec 2015 Jan 08;67(4):967-972. [doi: 10.1002/asi.23437]
59. Choudrie J, Banerjee S, Kotecha K, Walambe R, Karende H, Ameta J. Machine learning techniques and older adults processing of online information and misinformation: A covid 19 study. Comput Human Behav 2021 Jun;119:106716 [FREE Full text] [doi: 10.1016/j.chb.2021.106716] [Medline: 34866770]
60. Wilner AS. Cybersecurity and its discontents: Artificial intelligence, the Internet of Things, and digital misinformation. International Journal 2018 Jul 26;73(2):308-316. [doi: 10.1177/0020702018782496]
61. Guo Y, Hao Z, Zhao S, Gong J, Yang F. Artificial Intelligence in Health Care: Bibliometric Analysis. J Med Internet Res 2020 Jul 29;22(7):e18228 [FREE Full text] [doi: 10.2196/18228] [Medline: 32723713]

Abbreviations

AI: artificial intelligence
CPPs: citations per paper
MMR: measles, mumps, and rubella
WoS: Web of Science

Edited by A Mavragani; submitted 23.02.21; peer-reviewed by K Reuter, C Weiger, A Roundtree, JP Allem; comments to author 16.03.21; revised version received 30.03.21; accepted 20.12.21; published 25.01.22.

Please cite as:
Yeung AWK, Tosevska A, Klager E, Eibensteiner F, Tsagkaris C, Parvanov ED, Nawaz FA, Völkl-Kernstock S, Schaden E, Kletecka-Pulker M, Willschke H, Atanasov AG
Medical and Health-Related Misinformation on Social Media: Bibliometric Study of the Scientific Literature
J Med Internet Res 2022;24(1):e28152
URL: https://www.jmir.org/2022/1/e28152
doi:10.2196/28152
PMID:34951864

Original Paper

Using Large-scale Social Media Analytics to Understand Patient Perspectives About Urinary Tract Infections: Thematic Analysis

Gabriela Gonzalez[1], MD, MPH; Kristina Vaculik[2], MPH; Carine Khalil[2], PhD; Yuliya Zektser[3], MD; Corey Arnold[4], PhD; Christopher V Almario[2], MD, MSHPM; Brennan Spiegel[2], MD, MSHS; Jennifer Anger[5], MD, MPH

[1]Department of Urology, Davis School of Medicine, University of California, Davis, Sacramento, CA, United States

[2]Cedars-Sinai Center for Outcomes Research and Education, Los Angeles, CA, United States

[3]David Geffen School of Medicine, University of California, Los Angeles, CA, United States

[4]Computational Diagnostics, Departments of Radiology and Pathology, University of California, Los Angeles, CA, United States

[5]Department of Urology, University of California San Diego, La Jolla, CA, United States

Corresponding Author:
Gabriela Gonzalez, MD, MPH
Department of Urology
Davis School of Medicine
University of California, Davis
4860 Y Street, Suite 3500
Sacramento, CA, 95817
United States
Phone: 1 9167342222
Email: drggonzalez@ucdavis.edu

Abstract

Background: Current qualitative literature about the experiences of women dealing with urinary tract infections (UTIs) is limited to patients recruited from tertiary centers and medical clinics. However, traditional focus groups and interviews may limit what patients share. Using digital ethnography, we analyzed free-range conversations of an online community.

Objective: This study aimed to investigate and characterize the patient perspectives of women dealing with UTIs using digital ethnography.

Methods: A data-mining service was used to identify online posts. A thematic analysis was conducted on a subset of the identified posts. Additionally, a latent Dirichlet allocation (LDA) probabilistic topic modeling method was applied to review the entire data set using a semiautomatic approach. Each identified topic was generated as a discrete distribution over the words in the collection, which can be thought of as a word cloud. We also performed a thematic analysis of the word cloud topic model results.

Results: A total of 83,589 posts by 53,460 users from 859 websites were identified. Our hand-coding inductive analysis yielded the following 7 themes: quality-of-life impact, knowledge acquisition, support of the online community, health care utilization, risk factors and prevention, antibiotic treatment, and alternative therapies. Using the LDA topic model method, 105 themes were identified and consolidated into 9 categories. Of the LDA-derived themes, 25.7% (27/105) were related to online community support, and 22% (23/105) focused on UTI risk factors and prevention strategies.

Conclusions: Our large-scale social media analysis supports the importance and reproducibility of using online data to comprehend women's UTI experience. This inductive thematic analysis highlights patient behavior, self-empowerment, and online media utilization by women to address their health concerns in a safe, anonymous way.

(J Med Internet Res 2022;24(1):e26781) doi:10.2196/26781

KEYWORDS

female urology; urinary tract infections; health services research; social media; online community; online forum; latent Dirichlet allocation; data mining; digital ethnography

Introduction

Symptomatic acute bacterial cystitis, often used interchangeably with the term urinary tract infection (UTI), affects 60% of women once in their lifetime [1,2]. Based on self-reported data, the National Health and Nutrition Examination Survey identified a 12.6% annual incidence of UTI among women aged 18 years or older [3]. The cost of diagnosing and treating UTI has been estimated to be US $1.6 billion, which does not account for the management of recurrent UTI (rUTI) [1]. Up to 44% of women will have recurrent UTIs, defined as 2 or more UTIs within a 6-month period or 3 or more UTIs within 12 months [4].

Current qualitative studies among women with UTIs focus on prescription practice patterns, self-management strategies, and UTIs during pregnancy [5-7]. These studies are usually conducted in the clinical setting. Leveraging social media to understand the UTI experience of patients remains unexplored, despite the prominent use of online sources to supplement medical care among middle-aged to older adults [8,9].

To understand women's knowledge and experience with UTIs, we used digital ethnography to investigate patient perspectives via online media [9]. This research method adapts conventional ethnographic principles to understand a phenomenon in a population of interest by allowing investigators to study social media posts and conversations that serve as free-range, nonexperimental sources that may be generalized to a broader population of women with UTI [9]. Using online sources is a nonconventional tool to gather patient perspectives to better meet the medical needs of women with UTI seeking advice outside the clinical setting. Physicians should counsel patients based on their direct concerns. The National Institute of Diabetes and Digestive and Kidney Diseases–funded Prevention of Lower Urinary Tract Symptoms (PLUS) Network supported our efforts to complement its goals of understanding personal and environmental factors affecting women's bladder health [10]. The PLUS Consortium adopted the social ecological model, which considers interactions between social context and biology across the lifespan and views health behaviors as being determined by intrapersonal factors, interpersonal processes and primary groups, institutional factors, community factors, and public policy. We sought to characterize the awareness, patient experience, prevention strategies, and risk factors among women with UTI by conducting an ethnographic analysis of social media posts.

Methods

Data Acquisition

This study was found exempt by our institution's institutional review board. To gather large-scale online posts by women with UTI, we contracted with Treato, a web-based data-mining service company that utilizes extraction templates and a proprietary search algorithm designed to capture patient content. After consultation with the PLUS Consortium, our team chose a combination of keywords related to disease nomenclature, symptoms, treatment options, and exclusion terms to identify posts using their search algorithm (Multimedia Appendix 1). After extracting posts from online forums, we performed digital ethnography using qualitative thematic analysis and a latent Dirichlet allocation (LDA) topic modeling quantitative process that also facilitated qualitative results [11]. Combining both methods allowed us to ensure thematic saturation by analyzing the entire data set of identified posts.

Qualitative Analysis

After identifying posts that met our search criteria, the entire data set was randomized to ensure that we reviewed posts from various websites before reaching thematic saturation. We performed an inductive, iterative, open-coding, qualitative analysis until we could no longer identify unique themes. Two research team members were assigned to examine the extracted posts several times. The data were then organized into different units or codes, which provides sufficient detail for the reader even without the context. Therefore, the codes were supported by text fragments. This was an iterative process; hence, as unique inductive codes emerged, they were regrouped into more specific categories, and some were combined while others were placed in a superordinate category. Our goal was to avoid redundancy among the categories, so we created broad themes encompassing the categories.

LDA: A Quantitative and Qualitative Approach

To supplement the manual inductive coding process, we applied a second, more novel technique, LDA, that allowed for the review of the entire data set. LDA is an unsupervised probabilistic topic model process that relies on the contextual co-occurrence of words to identify patterns of words that, when found together, have a semantic meaning [11,12]. For example, the word "bank" can have different meanings when paired with "money" versus "water." This model generates outputs as topics that can be understood with the concept of a "word cloud," comprised of words that are ranked higher in a corresponding topic, if these co-occurred frequently in the social media posts. These topics were interpreted for thematic analysis. Each topic has an assigned prevalence value, which represents the quantity of words in the collection assigned to the topic divided by the total number of words. The word cloud topics were sorted based on their respective prevalence (quantitative signature) for review by the research team to identify a thematic interpretation. Consistency of theme allocation for a specific topic was confirmed by reviewing posts that contributed to the development of the word cloud. Table 1 demonstrates examples of the word cloud topics and their assigned themes. Combining both methods allowed for a comprehensive review of the results using a semiautomated approach and a manual inductive coding process to capture a broad understanding of the experience of patients with UTI.

Table 1. Example word cloud topics with their respective prevalence and assigned themes.

Topics	Prevalence (%)	Themes
uti, burning, symptoms, urination, urine, urinate, frequent, sensation, urinating, urge, ago, having, past, started, feel, time, discomfort, painful, year, week	15	Online community support (symptom sharing)
years, utis, recurrent, months, year, ago, antibiotics, infections, cause, menopause, sex, antibotics, colonized, time, month, bladder, mg, antibiotic, uti, stopped, tried, chronic	14	Recurrent UTI[a], risk factors and prevention
uti, taking, antibiotics, antibiotic, prescribed, mg, infection, took, macrobid, bactrim, taken, week, amoxicillin, started, medication, nitrofurantoin, got, pills, allergic, prescription	11.5	Antibiotics
use, clean, shower, utis, soap, uti, using, underwear, make, baths, used, cause, infections, toilet, utis, skin, sure, wipe, change, hot water, douche	10.5	Hygiene, risk factors and prevention
infection, uti, antibiotics, need, kidneys, bladder, cause, utis, infections, untreated, symptoms, treated, pregnancy, worse, sounds, checked, definitely, serious, asap, away	5	Treatment (untreated UTI)
bladder, cystitis, ic, symptoms, urologist, interstitial, years, flare, diagnosed, help, diet, chronic, uti, condition, infection, painful, urethra, thought, inflammation, pelvic, misdiagnosed	4	Diagnosis (overlap with interstitial cystitis)
cranberry, drink, juice, uti, drinking, help, lots, utis, make, helps, utis, sure, pills, fluids, antibiotics, flush, bladder, need, avoid, drinks	3.5	Alternative therapies

[a]UTI: urinary tract infection.

Results

We identified 83,589 posts written by 53,460 unique users found on 859 websites from January 2016 to December 2018.

LDA Topic Modeling Themes

We identified a total of 105 themes using LDA, which were grouped into 9 categories to avoid redundancy and provide an overview of the topics represented online (Table 2). Additionally, there was significant overlap with our hand-coding approach, so the data were synthesized into 7 themes with subthemes (Table 2) to represent results from both methods. Our hand-coding approach facilitated more descriptive interpretations.

Table 2. Nine categories of themes found using latent Dirichlet allocation.

Categories	Prevalence (N=105), n (%)
Online community support	27 (25.7)
Risk factors and prevention	23 (22.0)
Self-management strategies	14 (13.3)
Antibiotics	11 (10.5)
Alternative therapies	7 (6.7)
Access to care	7 (6.7)
Treatment options	7 (6.7)
Quality of life	6 (5.7)
Diagnosis	3 (2.9)

Inductive Thematic Analysis

Qualitative hand-coding analysis yielded 7 themes with subthemes related to the knowledge and experience of women with UTI symptoms (Table 3).

Table 3. Themes and illustrative examples from 200 online posts.

Theme	Subtheme	Example quotes
Quality of life	• Impact on sexual health • Fear • Pain • Self-blame	• "UTIs always dictate my life and make me feel very low, I'm always worried about going out for the day and whether there will be toilets around" • "feeling helpless, like it's a nightmare: rUTIs"
Knowledge acquisition	• Differential diagnosis • UTI[a] workup • Seeking an etiology • IC[b] symptom overlap • Untreated UTIs	• "I was told I had trace blood in urine…what's trace?... Dr. didn't explain." • "Some of what I thought was constant UTIs was actually interstitial cystitis - an inflammation triggered by acidic foods." • "Interstitial cystitis is similar to recurring UTI, with WBC and no bacteria"
Support of online community	• Online engagement • Seeking advice • Symptom sharing • Information exchange • Self-management strategies • Unique populations	• "Are these things true or is it a myth?" • "The information here to understand my dreadful cycles of UTIs is better than any library." • "Just wondering if anyone has ever gotten or heard of someone that has gotten misdiagnosed with recurrent urine infections when they have other bladder problems." • "Girl, get to the doctor ASAP. UTIs are no joke. They basically travel up the urinary tract and you can get a bladder infection or even a kidney infection. You don't want that. Go to urgent care if you have to. They just need a pee sample. Amoxicillin is your friend!"
Health care utilization	• Access barriers • Treatment location • Patient-physician interactions	• "I'm losing a lonely battle with the medical community as these UTIs won't go away. I'm glad you are here." • "I need to take an action list...no one gives me straight answers." • "If you can't go to the ER due to no insurance, what about a Planned Parenthood? When I was between insurance plans, I went there for UTI's, and all other lady things." • "Mismanagement always happens so recommend seeing a urogynecologist, not a GP."
Risk factors and prevention	• Dehydration • Hygiene • Anatomical differences • Hormonal imbalances • Gynecologic factors • Comorbid conditions	• "Women should be drinking 3L of water per day to prevent UTIs." • "I've heard that douching, tampons soaked in different things (tried in the past) can help treat and prevent urine infections." • "Prolapse after pregnancy and 'abnormal anatomy' making more prone to infections."
Antibiotic treatment	• Treatment duration • Safety and side effects • Multidrug-resistant bacteria • Effect on microbiome • Culture-directed antibiotic treatment	• "Prescribing antibiotics for UTIs seems weird." • "It [antibiotic] slayed me! In bed, exhaustion and major aches like I had the flu." • "I've done a 5-day course of antibiotics…no improvement" • "You get better improvement with antibiotics for 7 days with no sex."

Theme	Subtheme	Example quotes
Alternative therapies	• Homeopathy • Dietary modifications	• "Homeopathic medicine…is more tailored to you, they spend an hour talking to you and can treat your urine infection." • "D-mannose can be used as a treatment or preventatively and I think would be most useful for your mom since it works by sort of lubricating the urinary tract, so the bacteria are unable to stick on and cause inflammation." • "Infections damage bladder lining. recommend drinking baking soda."

[a]UTI: urinary tract infection.

[b]IC: interstitial cystitis.

Quality of Life

The first theme was the quality-of-life burden associated with UTI episodes. The impact on women's sexual health was frequently mentioned in the context of limiting intercourse due to aggravating symptoms (pain) and managing postcoital antibiotic use. Women described significant negative emotions and hopelessness when they sought self-management strategies and medical care. Self-blame was central to the negative emotions described, as women searched for inherent personal factors causing repetitive infections. Fear of worsening symptoms and progression to pyelonephritis was frequently mentioned.

Knowledge Acquisition

Patient knowledge acquisition was another major theme. Based on the keyword content of the posts, it appeared that women consulted online resources at different time intervals to supplement their decision-making while experiencing UTI symptoms or seeking medical care. Some users focused on identifying a differential diagnosis and a specific etiology, while others described self-blame. There was a lack of consensus regarding the optimal work-up and management of UTIs, as evidenced by people providing inconsistent advice to each other on these forums. The misdiagnosis of rUTIs and interstitial cystitis due to symptom overlap, delayed referral, and perceived lack of physician knowledge appeared frequently.

Online Community Support

The value and gratitude expressed for the support provided by online communities was another identified theme. In addition to the plethora of information exchanged, including symptom sharing and lay recommendations, we identified geriatric patients, pregnant women, and those with rUTIs as unique populations who frequently appeared as the subject matter of posts with special considerations. Pregnant women had specific interests regarding antibiotic safety and the development of pyelonephritis.

Health Care Utilization

The third theme was health care utilization with subthemes centering on the contextual factors influencing whether or not people sought care. Posts included concerns about minimal insurance coverage or being uninsured. Additionally, multiple medical visits for recurrent infections appeared to cause fatigue, frustration, and loss of work productivity. Furthermore, the perceived lack of illness clarity and lack of cure affected the way users commented about their experience.

Risk Factors and Prevention

Risk factors and prevention was another theme we identified. Women sought to understand their respective predisposing contributions due to day-to-day activities. The appropriate preventive hydration level was frequently mentioned, with various levels ranging from 1 to 3 L of water. Additionally, genital hygiene (self and partners') practices were discussed. Pelvic organ prolapse and vaginal atrophy were perceived to increase the risk of UTIs. Diabetes and dementia were also frequently mentioned risk factors. Gynecologic factors that were discussed included methods of contraception and menstrual cycle sanitation products.

Antibiotic Treatment and Alternative Therapies

Treatment of UTIs with antibiotics was another identified theme. The appropriate duration and variation in the prescribed length of treatment were discussed, as were the safety and side effects of antibiotics for pregnant and nonpregnant women. The online community misunderstood antibiotic resistance as a patient characteristic that developed, rather than as a bacterial phenomenon. Recommendations to restore the natural gut microbiome were exchanged. The final identified theme was alternative therapies beyond antibiotics to self-manage symptoms, ameliorate current infections, or prevent further UTIs. Some of these alternative therapies included bacteriophage therapy, cranberry products, d-mannose, vitamin C, probiotics, bladder instillation of hyaluronic acid, and oral activated charcoal treatments.

Discussion

Principal Results

Our ethnographic study of social media posts on UTIs revealed information on illness experience, lay knowledge, and concerns among women. Unlike prior qualitative studies, we presented patient perspectives that are likely more diverse and candid than data gathered from specialty clinics [5-7]. We found a strong online community support network created via forums to exchange information among peers, which may have partially resulted from frustrations and challenges with medical care. We found that UTIs cause a significant burden on daily activities and, as a result, women engage in supportive conversations

about physician interactions, antibiotics, alternative therapies, risk factors, and prevention strategies to bridge knowledge gaps and obtain reassurance from peers.

We captured broad and diverse patient experiences using two methods to conduct digital ethnography. However, our inductive hand coding provided more granular details as expected from directly analyzing quotes, which helped us comprehend online discussions and women's perspectives for specific themes. For example, culture-directed antibiotic treatment was a unique patient concern identified with our hand coding of posts that was not found using LDA. Although the LDA word clouds consistently represented quality-of-life concerns, hand coding provided examples of fears women faced. Half of the LDA themes related to community support and identifying risk factors and prevention strategies, which was consistent with our hand-coding results.

Comparison With Prior Work

Ghouri et al [13] previously conducted 15 telephone interviews with women with prior documented UTI and described feelings of hopelessness and lack of support in this group. Our study highlights the support experienced by patients with UTI who exchanged information online. Our findings support prior results that social media is an integral part of processing medical information and can facilitate patient engagement for the exchange of condition-specific knowledge [14,15]. Additionally, it has been widely documented that online forum discussions allow for better intake and information processing [16-18].

Our study was broad, capturing different populations of women. This better allowed the analysis to be guided by the direct, anonymous discussions of patients, making it more likely to be generalizable to the UTI population at large. Prior work only surveyed patients recruited from clinical settings, those with rUTI, and pregnant women [5-7]. Our work, on the other hand, captured the perspectives of several populations in a single analysis, including pregnant women, geriatric women, and women with rUTI across a collection of websites. Geriatric patients, pregnant women, older women, and women with rUTIs were frequently identified in posts with unique concerns, suggesting the need for more targeted outreach to these special populations.

Unlike prior online studies, our study design has the advantage of analyzing multiple websites [5,6]. Flower et al [5] conducted an analysis of 1 online self-help forum for patients with rUTI to understand how women manage their rUTIs. Our findings of alternative therapies, antibiotic concerns, and patient-physician interactions were similar. However, due to the broader sample size, we found other alternative therapies not previously described, additional antibiotic concerns (eg, treatment duration, human microbiome, bacterial resistance, and culture-guided treatment), and more complex health care barriers (eg, specialty care access, insurance coverage, and presumed level of care required for treatment). To our knowledge, the theme of risk factors and prevention has not been previously described in prior UTI online forum literature. Additionally, our findings that pregnant women were concerned about the progression of cystitis to pyelonephritis, as well as the effect of antibiotic use (or lack of treatment) on fetal development, were consistent with those previously found in a study analyzing online content to understand UTIs and antibiotic use in the pregnant population [6].

The semistructured interview style of many qualitative studies may limit and potentially narrow the scope of what patients share in clinical settings. One qualitative one-on-one interview study of 21 women recruited from a larger primary care trial found that patients wanted clinicians to address quality-of-life impact and that they were receptive to the strategy of antibiotic delay, which allows for 48 hours to reassess if infection symptoms subside before starting antibiotics [7]. Women who enrolled in this randomized study were, by default, receptive to different management strategies. Although we found similar quality-of-life concerns, we also identified self-blame, mismanagement in the primary care setting, delayed referrals, inconsistent counseling about treatment guidelines, unmet expectations, and the practice of culture-directed antibiotic use, which were not previously characterized.

Online discussion points were in agreement with the 2019 American Urological Association's guidelines for uncomplicated rUTI, which recommend first-line antibiotic agents and promote culture-directed antibiotic treatments rather than empiric treatment [19]. This suggests that many patients with UTIs were well educated on the topic. However, there were inconsistencies discussed online for the role of cystoscopy and upper tract imaging, despite the recommendations provided in the updated guidelines to avoid those diagnostic studies in uncomplicated cases [19]. This may have been due to the fact that our data were collected before the publication of the updated guidelines. Concerns about antibiotic collateral damage mentioned in the literature and discussed in the rUTI guidelines were also supported by our analysis [19].

Limitations

Despite the innovation and patient inclusivity of our study, there are important limitations that can inform future work. We did not have access to demographic information, and the website content could have been restricted by the sample of websites accessed by Treato and our search strategy. Although we focused our analysis on women, it is possible that some men participated on the forums and were included in the analyses. The anonymous data may also contribute to patient misclassification since we cannot confirm a diagnosis, but our best attempt was made using contextual factors. Additionally, our analysis and conclusion relied on the degree to which individuals post online. We could not characterize specific posts' engagement level, such as individual read and reply counts. Our study, by default, excluded those women who do not exchange medical information on the internet.

Conclusions

Digital ethnography combining qualitative analysis and LDA allowed us to analyze free-range patient perspectives, which are currently not found in the UTI literature. First, unlike focus group studies, anonymity is a clear driver of candid, honest conversations, facilitating online users to provide support and address the most important concerns. Second, there was a pervasive element of fear: fear of not treating UTIs, as well as

fear of the sequelae associated with antibiotic treatments. Finally, the use of online forums empowered women to self-manage their condition and take their care into their own hands. Our findings also demonstrate the reliability of using online social media data to learn about patient behavior and decision-making, which is important to guide how we engage with patients and disseminate society-sponsored guidelines. Patient information, outreach, and treatment guidelines by medical societies must be congruent with patients' concerns. Physicians can use this data to discuss misconceptions and improve patient-centered care.

Acknowledgments

This study was funded by a pilot grant from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium (to authors JA, BS).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Search and exclusion terms used in the Treato algorithm.
[PNG File , 2146 KB - jmir_v24i1e26781_app1.png]

References

1. Foxman B. Epidemiology of urinary tract infections: incidence, morbidity, and economic costs. Dis Mon 2003 Feb;49(2):53-70. [doi: 10.1067/mda.2003.7] [Medline: 12601337]
2. Foxman B. The epidemiology of urinary tract infection. Nat Rev Urol 2010 Dec;7(12):653-660. [doi: 10.1038/nrurol.2010.190] [Medline: 21139641]
3. Suskind AM, Saigal CS, Hanley JM, Lai J, Setodji CM, Clemens JQ, Urologic Diseases of America Project. Incidence and Management of Uncomplicated Recurrent Urinary Tract Infections in a National Sample of Women in the United States. Urology 2016 Apr;90:50-55 [FREE Full text] [doi: 10.1016/j.urology.2015.11.051] [Medline: 26825489]
4. Brubaker L, Carberry C, Nardos R, Carter-Brooks C, Lowder JL. American Urogynecologic Society Best-Practice Statement. Female Pelvic Med Reconstr Surg 2018;24(5):321-335. [doi: 10.1097/spv.0000000000000550]
5. Flower A, Bishop FL, Lewith G. How women manage recurrent urinary tract infections: an analysis of postings on a popular web forum. BMC Fam Pract 2014 Sep 26;15(1):162 [FREE Full text] [doi: 10.1186/1471-2296-15-162] [Medline: 25260870]
6. Ghouri F, Hollywood A, Ryan K. Urinary tract infections and antibiotic use in pregnancy - qualitative analysis of online forum content. BMC Pregnancy Childbirth 2019 Aug 13;19(1):289 [FREE Full text] [doi: 10.1186/s12884-019-2451-z] [Medline: 31409404]
7. Leydon GM, Turner S, Smith H, Little P. The journey from self-care to GP care: a qualitative interview study of women presenting with symptoms of urinary tract infection. Br J Gen Pract 2009 Jul 01;59(564):e219-e225. [doi: 10.3399/bjgp09x453459]
8. Social Media Fact Sheet. Pew Research Center. 2019 Jun 12. URL: https://www.pewresearch.org/internet/fact-sheet/social-media/ [accessed 2020-12-01]
9. Garcia AC, Standlee AI, Bechkoff J, Yan Cui. Ethnographic Approaches to the Internet and Computer-Mediated Communication. J Contemp Ethnogr 2009 Feb 01;38(1):52-84. [doi: 10.1177/0891241607310839]
10. Harlow BL, Bavendam TG, Palmer MH, Brubaker L, Burgio KL, Lukacz ES, et al. The Prevention of Lower Urinary Tract Symptoms (PLUS) Research Consortium: A Transdisciplinary Approach Toward Promoting Bladder Health and Preventing Lower Urinary Tract Symptoms in Women Across the Life Course. J Womens Health (Larchmt) 2018 Mar;27(3):283-289 [FREE Full text] [doi: 10.1089/jwh.2017.6566] [Medline: 29634445]
11. Campbell JC, Hindle A, Stroulia E. Chapter 6 - Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data. In: Bird C, Menzies T, Zimmermann T, editors. The Art and Science of Analyzing Software Data. Boston, MA: Morgan Kaufmann; Jan 01, 2015:160.
12. Dzubur E, Khalil C, Almario CV, Noah B, Minhas D, Ishimori M, et al. Patient Concerns and Perceptions Regarding Biologic Therapies in Ankylosing Spondylitis: Insights From a Large-Scale Survey of Social Media Platforms. Arthritis Care Res (Hoboken) 2019 Feb 29;71(2):323-330 [FREE Full text] [doi: 10.1002/acr.23600] [Medline: 29781587]
13. Ghouri F, Hollywood A, Ryan K. 'There is no choice apart from antibiotics…': Qualitative analysis of views on urinary infections in pregnancy and antimicrobial resistance. Health Expect 2020 Jun 29;23(3):644-650 [FREE Full text] [doi: 10.1111/hex.13044] [Medline: 32113189]
14. Gonzalez G, Vaculik K, Khalil C, Zektser Y, Arnold C, Almario CV, et al. Using Digital Ethnography to Understand the Experience of Women With Pelvic Organ Prolapse. Female Pelvic Med Reconstr Surg 2020 Sep 9;27(2):e363-e367. [doi: 10.1097/spv.0000000000000938]

15. Van Uden-Kraan CF, Drossaert CHC, Taal E, Shaw BR, Seydel ER, van de Laar MAFJ. Empowering processes and outcomes of participation in online support groups for patients with breast cancer, arthritis, or fibromyalgia. Qual Health Res 2008 Mar;18(3):405-417. [doi: 10.1177/1049732307313429] [Medline: 18235163]
16. Allen C, Vassilev I, Kennedy A, Rogers A. Long-Term Condition Self-Management Support in Online Communities: A Meta-Synthesis of Qualitative Papers. J Med Internet Res 2016 Mar 10;18(3):e61 [FREE Full text] [doi: 10.2196/jmir.5260] [Medline: 26965990]
17. Cole J, Watkins C, Kleine D. Health Advice from Internet Discussion Forums: How Bad Is Dangerous? J Med Internet Res 2016 Jan 06;18(1):e4 [FREE Full text] [doi: 10.2196/jmir.5051] [Medline: 26740148]
18. Thomas M. Learning within incoherent structures: the space of online discussion forums. J Comput Assist Lear 2002 Jan 01;18:351-366. [doi: 10.1046/j.0266-4909.2002.03800.x]
19. Anger J, Lee U, Ackerman AL, Chou R, Chughtai B, Clemens JQ, et al. Recurrent Uncomplicated Urinary Tract Infections in Women: AUA/CUA/SUFU Guideline. J Urol 2019 Aug;202(2):282-289. [doi: 10.1097/ju.0000000000000296]

Abbreviations

LDA: latent Dirichlet allocation
PLUS: Prevention of Lower Urinary Tract Symptoms
rUTI: recurrent urinary tract infection
UTI: urinary tract infection

Edited by A Mavragani; submitted 28.12.20; peer-reviewed by M Holter, S Uddin; comments to author 18.05.21; revised version received 23.10.21; accepted 10.12.21; published 25.01.22.

Please cite as:
Gonzalez G, Vaculik K, Khalil C, Zektser Y, Arnold C, Almario CV, Spiegel B, Anger J
Using Large-scale Social Media Analytics to Understand Patient Perspectives About Urinary Tract Infections: Thematic Analysis
J Med Internet Res 2022;24(1):e26781
URL: https://www.jmir.org/2022/1/e26781
doi:10.2196/26781
PMID:35076404

Original Paper

An Exploration of e-Cigarette–Related Search Items on YouTube: Network Analysis

Hassan Dashtian[1*], PhD; Dhiraj Murthy[1*], PhD; Grace Kong[2*], PhD

[1]The Computational Media Lab and School of Journalism and Media, The University of Texas at Austin, Austin, TX, United States

[2]The Department of Psychiatry at Yale School of Medicine, New Haven, CT, United States

[*]all authors contributed equally

Corresponding Author:
Dhiraj Murthy, PhD
The Computational Media Lab and School of Journalism and Media
The University of Texas at Austin
300 W. Dean Keeton (A0900)
Austin, TX, 78712-1069
United States
Phone: 1 512 471 5775
Email: Dhiraj.Murthy@austin.utexas.edu

Abstract

Background: e-Cigarette use among youth is high, which may be due in part to pro–e-cigarette content on social media such as YouTube. YouTube is also a valuable resource for learning about e-cigarette use, trends, marketing, and e-cigarette user perceptions. However, there is a lack of understanding on how similar e-cigarette–related search items result in similar or relatively mutually exclusive search results. This study uses novel methods to evaluate the relationship between e-cigarette–related search items and results.

Objective: The aim of this study is to apply network modeling and rule-based classification to characterize the relationships between e-cigarette–related search items on YouTube and gauge the level of importance of each search item as part of an e-cigarette information network on YouTube.

Methods: We used 16 fictitious YouTube profiles to retrieve 4201 distinct videos from 18 keywords related to e-cigarettes. We used network modeling to represent the relationships between the search items. Moreover, we developed a rule-based classification approach to classify videos. We used betweenness centrality (BC) and correlations between nodes (ie, search items) to help us gain knowledge of the underlying structure of the information network.

Results: By modeling search items and videos as a network, we observed that broad search items such as *e-cig* had the most connections to other search items, and specific search items such as *cigalike* had the least connections. Search items with similar words (eg, *vape* and *vaping*) and search items with similar meaning (eg, *e-liquid* and *e-juice*) yielded a high degree of connectedness. We also found that each node had 18 (SD 34.8) connections (common videos) on average. BC indicated that general search items such as *electronic cigarette* and *vaping* had high importance in the network (BC=0.00836). Our rule-based classification sorted videos into four categories: e-cigarette devices (34%-57%), cannabis vaping (16%-28%), e-liquid (14%-37%), and *other* (8%-22%).

Conclusions: Our findings indicate that search items on YouTube have unique relationships that vary in strength and importance. Our methods can not only be used to successfully identify the important, overlapping, and unique e-cigarette–related search items but also help determine which search items are more likely to act as a gateway to e-cigarette–related content.

(J Med Internet Res 2022;24(1):e30679) doi:10.2196/30679

KEYWORDS

electronic nicotine delivery systems; vaping; social media; search engine; natural language processing; social network analysis

Introduction

Background

e-Cigarette use has grown exponentially since its introduction to the US market in 2007. Currently, e-cigarettes are the second most used tobacco product (4.5%), following cigarettes (14%), among adults [1]. The highest use is among young adults aged 18 to 24 years (9.3%) [1]. A concerning trend is that US adolescents have been using e-cigarettes at an alarmingly high rate; 19.6% of high school students reported using e-cigarettes in the past month (the cigarette smoking rate is 4.6%), and 38.9% of the past-month users were using e-cigarettes for ≥20 days [2]. Similar trends in e-cigarette use have been reported in other countries. For example, e-cigarette use ranges from 0.2% to 27% in European countries [3]. A survey conducted by Wang et al [4] found that a quarter of young Chinese adults had used e-cigarettes. Given the high rates of e-cigarette use worldwide, particularly among young people, it is imperative to understand the appeal, use patterns, and marketing associated with the use of e-cigarettes to inform prevention strategies.

YouTube and e-Cigarettes

Analysis of social media platforms such as YouTube, which are widely used by youth [5] and have also been shown to have high e-cigarette content [6], could provide insights into e-cigarette use trends and marketing strategies that youth are exposed to. YouTube is a free web-based video streaming service that allows users to view and upload videos, post comments, and rate videos. Users can also interact by subscribing to each other's YouTube channels and sharing opinions on the video content through writing comments on videos and *liking* or *disliking* videos. YouTube has 2 billion users, which is a third of all internet users, and people currently spend >1 billion hours per day watching web-based videos on the platform [7]. YouTube, along with Facebook, continues to dominate web-based social media use in 2021. Although social media platforms, most notably Instagram, Snapchat, and TikTok, have been very popular among youth, they also actively use YouTube [5].

A concerning finding is that pro–e-cigarette content is readily available on YouTube [6]. An examination of e-cigarette–related YouTube videos from 2012 to 2013 identified 28,000 videos, and these videos were viewed >100 million times, indicating a high level of content consumption [6]. Prior research has also identified that 85% of the pro–e-cigarette content on YouTube is by the e-cigarette industry [8], and the marketing content includes comparisons with cigarettes and emphasizes themes that e-cigarettes are cleaner and cheaper than combustible cigarettes and that they can be used anywhere [9]. Other content areas on YouTube include characterizations of a variety of use behaviors such as vape tricks, which involve using e-cigarettes to blow large, thick amounts of exhaled aerosol (ie, clouds) or shapes [10], and unorthodox use of e-cigarettes, which is manipulating the product to be used for unintended purposes [11,12]. YouTube videos have also been used to understand the way users puff e-cigarettes [13], product characteristics through product reviews [14], the presence of nicotine warning labels [15], and health information regarding e-cigarettes [16], as well as identify and characterize e-cigarette users [10].

Search Algorithms on YouTube

As YouTube is being used widely to understand e-cigarette–related behaviors, it is important to advance the methods used to obtain and analyze videos. Researchers have searched YouTube for keywords of interest to obtain information on e-cigarettes. Most existing studies that examine e-cigarettes on YouTube have used a range of general search items, such as *electronic cigarette*, *e-cigarette*, *e-cig*, *vaping*, *vape*, *e-liquid*, and *e-juice* [8-10,14,15]. Less known is how well these search items yield videos relevant to e-cigarettes and whether these search items yield comparable or unique results. Currently, there is no guidance for researchers to evaluate how related search items function to yield relevant and parsimonious results on YouTube. Such evaluation is necessary as search items drive the content being shown, which can ultimately shape conclusions reached to inform policy regarding e-cigarettes. Thus, a goal of our study is to use novel methods (eg, network modeling and rule-based classification) to evaluate the association between similar e-cigarette–related search items and search results, as well as to identify which search items act as gateways to e-cigarette–related content.

Both the search and recommendation algorithms on YouTube were structured from the platform's infancy to drive views of videos [17]. YouTube recommends a series of related videos to users in response to the video currently being played by the users or to a specific search item being used in the search engine. YouTube's search and recommendation algorithms use machine-learned approaches and are predominantly based on the users' cumulative viewing experiences. As YouTube's algorithm is proprietary, the first step in the characterization of e-cigarette–related information on YouTube is to reconstruct the resulting associations between the content of the videos related to this topic that are retrieved from querying YouTube's search engine. We applied network modeling to study the *relatedness* of search items and the associations between the content of the videos retrieved from different search items. This approach has been successfully used in other studies. Abul-Fottouh et al [18] modeled vaccine-related videos on YouTube as a network to evaluate how the platform recommends vaccination-related videos to its users. Murthy [19] used features of YouTube videos to illustrate and visualize variations in video recommendations based on the language of a video. However, in this study, we used network modeling instead of YouTube's recommendation system to model YouTube search engine results. Network modeling will help researchers understand how different e-cigarette search items are related in YouTube search results.

We have also used the features of the networks, particularly betweenness centrality (BC) [20], which measures the importance of entities in a network, which have a positional advantage in that they connect the shortest (geodesic) paths between other pairs of entities. A practical application of BC is to determine what search items in the network of information of e-cigarette–related videos are more central and thus provide information seekers with more relevant information in the

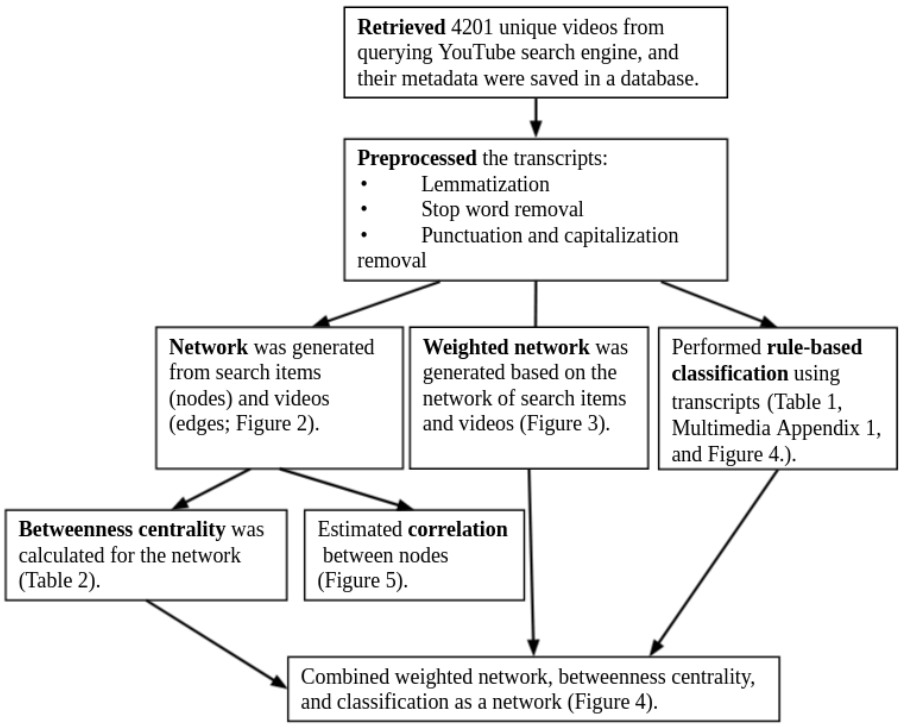
YouTube search engine. We used correlations between nodes (ie, search items) to characterize the network to identify the relationships between e-cigarette–related search items and the level of importance of each search item in the information network of these search items. Then, we used a rule-based classification to assess whether our search items correctly identified videos related to e-cigarettes.

In summary, the aim of this study is to use network modeling and rule-based text classification methods to understand how the search for e-cigarettes on YouTube is affected by differences in search items. These methods can be used to identify the overlap or uniqueness of search items related to e-cigarettes on YouTube. Assessing whether the search items are parsimonious and whether redundancies can be reduced could inform future work by assisting others in optimizing their search queries to retrieve and analyze information related to e-cigarettes on YouTube videos.

Methods

Figure 1 illustrates the framework we used to collect, process, and analyze YouTube data.

Figure 1. Project architecture; relevant figures and tables are referenced in parentheses.



Video Identification

First, we created fictitious viewer accounts to simulate individuals searching for e-cigarette–related content on YouTube. A separate mobile SIM card was obtained to create a Google account for each fictitious account. The fictitious YouTube profiles comprised African American women and men aged 16 and 24 years (4/16, 25% in total), Hispanic women and men aged 16 and 24 years (4/16, 25% in total), and 2 sets of White women and men aged 16 and 24 years (8/16, 50% in total). We used common African American, Hispanic, and White first and last names to create the fictitious profiles. We created more White profiles to better reflect the e-cigarette use populations [21]. With each of the 16 fictitious viewer profiles, we used a factory-reset Android phone with Orbot (The Tor Project) [22], an app that allows the use of the anonymized Tor IP bridge with any app, to search YouTube for the following keywords related to e-cigarettes in July 2020: *e-cigarette*, *e-cig*, *electronic cigarette*, *e-liquid*, *ENDS*, *e-juice*, *vape*, *vaping*, *vape juice*, *box mods*, *cigalikes*, *disposable e-cigs*, *disposables*, *disposable vape*, *pod mods*, *vape mods*, *vape pens*, and *vape pods*. We factory reset the phone and Tor IP address after collecting data for each user so that any personalization by YouTube was erased. Other researchers have used general search items related to e-cigarettes on YouTube to identify trends in perceptions, use, marketing and sales related to e-cigarettes [8-10,14,15]. On the basis of this prior literature, we chose a single general term related to *e-cigarettes* as our search item to assess whether our methods could be used to evaluate the associations between the results derived from these search items.

Across all search items and for each profile, we requested 140 videos (7 pages with 20 videos on each page) and collected a total of 5875 videos after removing the duplicates across different profiles (see Multimedia Appendix 1 for a breakdown of the collected videos by search item after the duplicates across the profiles were removed). Of those 5875 videos, we further removed 1674 (28.49%) duplicates across different search items, which resulted in our final data set of 4201 (71.51%) unique videos. We did not exclude duplicate videos when generating the network; we merely removed duplicates for classification purposes. For these 4201 videos, we extracted video transcripts directly from the YouTube application programming interface (API). We did not restrict the date of video upload as some

videos that were uploaded years ago could still be relevant. The video metadata that we requested from the YouTube API were public; therefore, the Yale Institutional Review Board deemed that institutional review board approval was not required.

Search Items as a Network

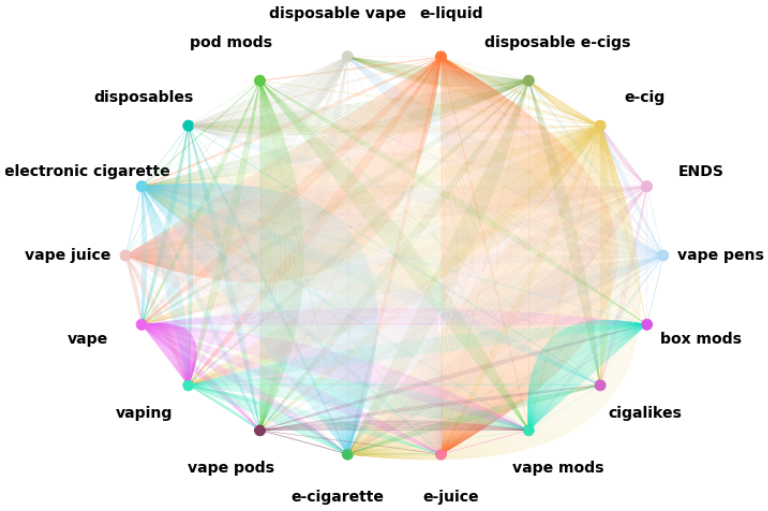
We modeled the resulting videos derived from the search items as a network. We assumed that each search item was a node and each video was an edge. If 2 search items had a connected edge, there was a common video in the resulting video data set. This method helped us to better visualize the connectivity of the search items and visualize the connectivity between search items. We then used social network analysis to study how search queries may be connected to each other. Connections between nodes in a network were a good proxy for the similarity between nodes. A search item's similarity with others could be measured by comparing the video IDs that we retrieved from querying YouTube.

BC Analysis

We then identified the nodes that had the most influence on the network of search items using BC, which measures the importance of entities (ie, nodes) in a network, which have a positional advantage in that they connect the shortest (geodesic) paths between other pairs of entities. This metric estimates all the shortest paths between every pair of search items in the network (Figure 2) and then computes the number of times a search item (node) is on the shortest path between 2 other nodes. Nodes with high BC may have considerable influence within a network by virtue of their control over information passing between others. They are also the ones whose removal from the network will most disrupt communications between other nodes as they lie on the largest number of paths connecting nodes.

Ultimately, BC indicates the importance of each node in a network. The BC of a node v is the sum—taken over all pairs (s, t) of source–target nodes—of the ratio between the number of shortest paths from s to t passing through v and the total number of shortest paths from s to t. Therefore, nodes with high BC belong to several shortest paths, whereas nodes with low BC belong to few shortest paths. A node with high BC is presumed to be a significant and influential node. An important feature of BC is that it is the measure of the degree to which a node is *between* other nodes in a graph or network, meaning that this node could act as a mediator in the network regardless of the frequency of connectivity [23,24]. This is very important in the context of YouTube as the platform's recommendation algorithms may use BC as a parameter to recommend videos based on the search items.

Figure 2. The network of search items and videos from our YouTube searches. Connections (edges) between the nodes (search items) indicate when the same videos resulted from 2 different search items, and a darker shaded edge line indicates a higher frequency of common results. ENDS: electronic nicotine delivery system.



Correlogram

We calculated the correlation coefficient between the nodes (search items) and the edges (number of videos) that linked to each other in the network and visualized the correlation using a correlogram. A correlogram can be used to visualize which nodes behave similarly (positive correlations) or differently (negative correlations).

Rule-Based Classification

Our methodological approach to classifying the themes of the videos used (1) a custom-developed Python script to extract video transcripts and video metadata, including number of views, title, category on YouTube, number of likes or dislikes, and comments; (2) preprocessing and lemmatization; (3) natural language processing (NLP) methods to extract keywords from video transcripts for their representation while removing stop words from the transcript; and (4) a rule-based classifier to categorize themes.

Preprocessing Data and Lemmatization

We cleaned the raw data by checking the quality of the raw text. Our preprocessing pipeline involved converting the raw text from the video transcripts into a vectorized form that was easily interpretable by our models. We used the spaCy (Explosion AI) library [25], an open-source NLP software library that comes with pretrained word vectors, to preprocess these transcripts. We used the provided *GloVe* [26] vectors from the *en_core_web_lg* model, which were trained on a variety of internet corpora. The use of word vectors allowed us to provide

our models with a meaningful representation in a vectorized form that encoded semantic similarity between words. Ideally, these word vectors would allow our models to use semantic relationships to determine whether the text we were classifying fell within a given theme, such as *e-cigarette device*. We used spaCy's list of 312 stop words (eg, *and*, *a*, *our*, and *my*) as they did not significantly contribute to the understanding of our data. We also lemmatized the words so that different word conjugations appeared identical to the classifier.

Rule Set

Approaches for the classification of text included rule-based methods, machine learning–based methods, and hybrid approaches that used both machine learning and rule-based methods [27,28]. We used rule-based methods to classify the text of the transcript of the video to identify whether the search items were indeed identifying e-cigarette videos. Rule-based approaches classify text into classes using a set of custom linguistic rules. These rules direct the algorithm to use semantically relevant elements of a text to identify similar categories based on the content of the texts and documents. Our rule set is detailed in Textbox 1. Indicators for each class were chosen by identifying common words in the video titles, descriptions, and transcripts.

We initially implemented a framework with >4 categories but discovered that some categories had very few cases. Therefore, we grouped categories with low frequencies into *other e-cigarette videos*. This category included videos that did not feature a specific e-cigarette device and were related to health information and news clips that contained news related to e-cigarettes. The other three categories included e-cigarette products featured on videos presented to YouTube users: specific e-cigarette devices that are used for vaping, the liquid that is used in the devices, and cannabis vaping products (Textbox 1). Our rationale for a rule-based method stems from 2 key reasons. First, we use a small number of classes (n=4), and these classes can be easily separated with a rule-based approach, and we do not need to train more sophisticated models to classify the video texts. Related text—specifically video transcripts—usually contain high dimensional nondiscriminative (irrelevant and noisy) words that can result in high computational costs and poor learning performance [29].

Textbox 1. Rule set for theme-based classification of e-cigarette–related videos.

Class type and indicators (videos can be categorized into >1 category)
e-Cigarette device
e-cigarette, electronic cigarette, e-cig, ecig, vape pen, vapepen, vape, pod, pod mod, disposable ecig, disposable vape, disposable, cigalike, box mod, boxmod, juul, puff bar, cartomizer, drip, drip tip, vape kit, mod, electronic nicotine delivery system, ENDS, puffbar, puff bar, sourin, blu, smok, leaf, markten, nicotek, vuse, fin, v2, 21st, atomizer, RDA, RTA, cartridge, ohm, wattage, watt, drip, cartomizer, ecig+device, vape+device
e-Liquid
e-juice, e-liquid, ejuice, eliquid, vapejuice, vape liquid, zamplebox
Cannabis vaping product**s**
cheeba, dab, firefly, ganja, gpen, hemp, indica, kush, marijuana, pax, pot, reefer, sativa, snoop dogg, weed, THC, cannabis, hash, wax, CBD
Other e-cigarette videos
These videos included e-cigarette–related videos but did not feature a specific e-cigarette device (eg, news clips and health information videos). We categorized videos that did not fit into any of the classes above into this category.

Results

Search Item Network

Figure 2 illustrates the network of 18 search results. In this network, each node (search item) is represented by a unique color, and darker shaded edges indicate a higher frequency of connections. Each line represents a common video (edge) between 2 nodes, indicating that a common video has been found in the search results for both search items. As expected, similar search items had more common videos. For example, *e-cigarette* and *electronic cigarette* had more common edges than *e-juice* and *disposable*.

Another way of representing the network of videos is to convert the number of connections (ie, edges) to a weight within the network for search items (Figure 3). We assumed that the nodes were connected through weighted edges. We counted the edges (common videos) between 2 nodes (search items) and normalized this number to a weight between 0 and 1, with 1 corresponding to the maximum number of edges between 2 nodes in the network and 0 indicating the minimum number of edges. The thickness of the edges is an indication of the weight, and thicker edges indicate more common videos and a stronger information flow and relationship between the 2 nodes.

From the networks illustrated in Figures 2 and 3, search items with similar words (eg, *vape* and *vaping*) and search items with similar meanings (eg, *e-liquid* and *e-juice*) have strong weighted edges, indicating that YouTube's search algorithm gave similar search results for these 2 categories. *e-Juice* and *vape juice* had 188 common edges, which was the maximum number of connections between 2 search items. *e-Juice* and *e-liquid* had 179 common edges, which was the second-largest number of connections. *Cigalike* had the lowest overall connections (n=82), whereas *e-cig* had the highest total connections (n=505). Search items *vape pens*, *ENDS*, and *e-cig* had at least one connected edge to other search items. *e-Cig* had the most connected edges to others (mean 29.7, SD 45.2), and *e-cigarette* had the second-highest connected edges (mean 27.6, SD 52.9). *Cigalikes* and *pod mods* had 4.8 and 6.6 common edges with other search items, respectively, which were the lowest connected edges. On

average, each node had 18.0 (SD 34.8) connections, indicating that there was a large variation in the number of connections between search items. This large SD could also be because of the fact that some of the search items were not connected (ie, there were no common videos between them). For example, *cigalikes* and *box mods* or *disposable vape* and *pod mods* had no connections. The complete list of the average connections is detailed in Table 1.

Figure 3. The network of search items and videos from our YouTube searches with weighted edges. The weighted edges, which represent the relative frequency of connections between 2 nodes, are represented by the thickness of each line. ENDS: electronic nicotine delivery system.

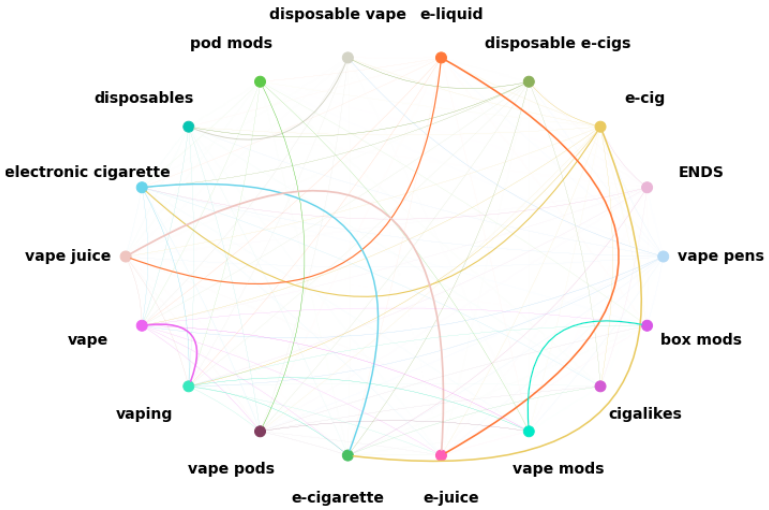


Table 1. Average number of connections (videos) for each node (search item) in the network and BC[a] of each node in the network.

Search item	Number of connections, mean (SD)	BC
e-cig	29.7 (45.2)	0.00836
e-cigarette	27.6 (52.9)	0.00836
vaping	26.2 (39.6)	0.00836
electronic cigarette	25.4 (45.1)	0.00836
e-juice	25.2 (59.7)	0.00232
vape	23.9 (39.6)	0.00407
vape juice	23.8 (52.3)	0.00407
e-liquid	21.5 (50.9)	0.00407
vape mods	15.2 (32.5)	0.00726
disposable vape	13.3 (24.8)	0.00049
disposables	11.5 (24.1)	0.00521
disposable e-cigs	11.5 (13.0)	0.00396
box mods	11.4 (32.4)	0.00056
vape pens	10.5 (9.3)	0.00836
vape pods	10.2 (11.5)	0.00836
ENDS[b]	7.9 (9.4)	0.00836
pod mods	6.6 (12.5)	0.00392
cigalikes	4.8 (6.4)	0.00105

[a]BC: betweenness centrality.

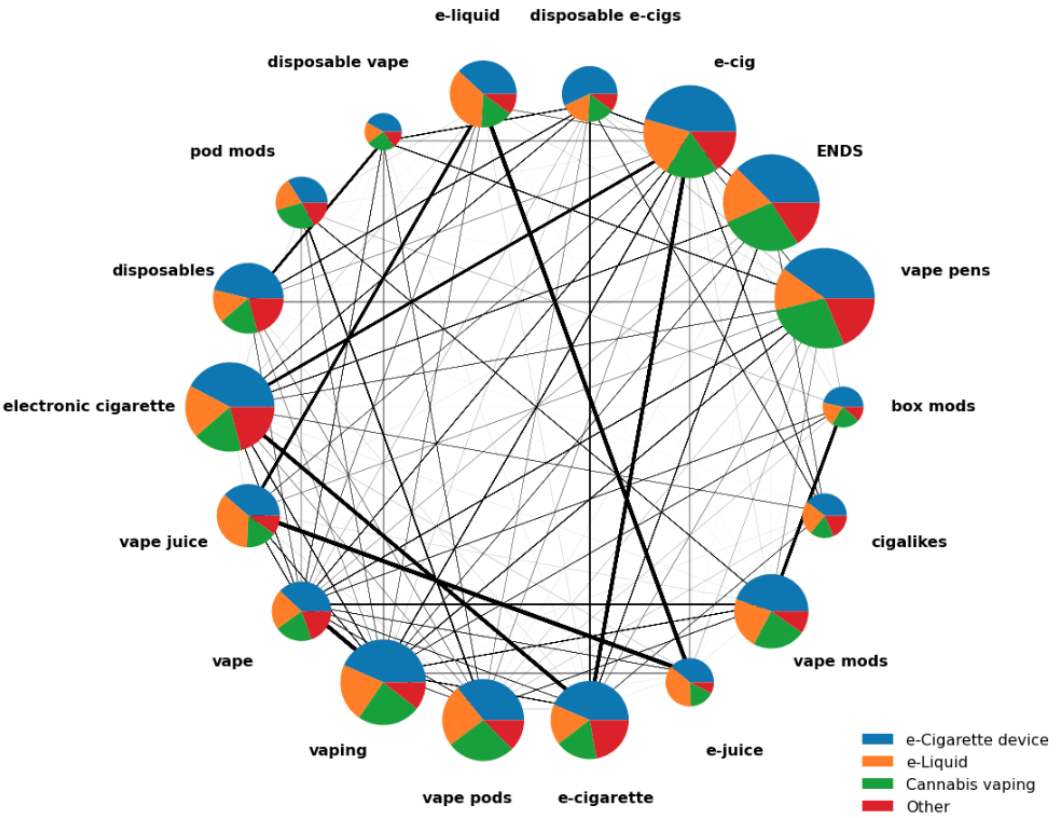
[b]ENDS: electronic nicotine delivery system.

BC and Rule-Based Classification

BC can be used to demonstrate the flow of information and provide a connective fabric between different search items in the network of e-cigarette–related videos. In our study, BC detected the dominant nodes (search items) that occupied a set of intermediate items observed between 2 other search items in the information network constructed with search items and YouTube videos. Generally, nodes with high BC are assumed to have a considerable impact on a network and can be

considered *important* nodes as they control the information flow between other nodes. Furthermore, a node with high BC may not have a high connectivity degree. For example, in our network, the *ENDS* node had 7.9 (SD 9.4) connections on average and was considered low relative to other search items; however, its BC was relatively high (Table 1 and Figure 4). High-BC nodes (with a low connectivity degree) can act as bridges between different clusters in a network (ie, subnetworks). Our network-based results indicate that there were subnetworks in the information network of e-cigarette–related videos, with search items such as *ENDS* acting as a bridge between these subnetworks. From a practical perspective, this means that users searching for ENDS could be exposed to ENDS content as well as videos in the *cigalike* and *e-liquid* subnetworks (eg, through YouTube's recommended videos feature), whereas searches for just *cigalike* or *e-liquid* would be more restricted in terms of both search results and videos being recommended.

Figure 4. Weighted network of search items. The thickness of an edge (thickness of the lines) indicates the weighted number of common videos between 2 search items. Betweenness centrality (BC) and classes (themes) of videos in search items are also shown. The size of each node (represented as a pie chart) indicates a node's BC, with a larger size indicating higher BC than a smaller size. Each pie chart shows the percentage of 4 classes in each search item. ENDS: electronic nicotine delivery system.

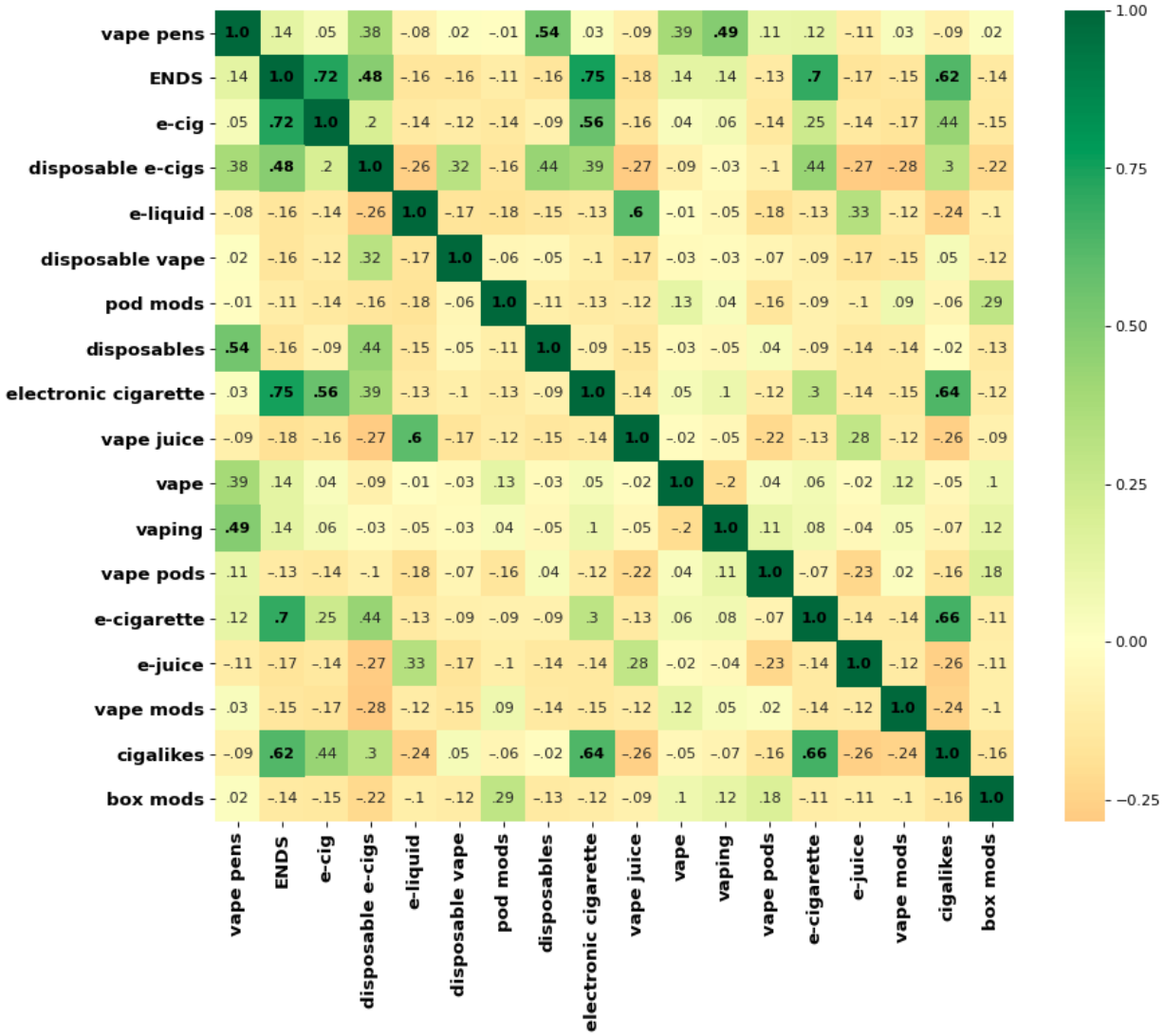


In Figure 4, the size of the pie indicates the relative BC values for the network of search items. For instance, a larger pie size indicates greater BC, suggesting that these are important nodes that serve as gateways to most videos in the network of videos derived from our data set. Nodes that did not show high levels of importance are represented by midsize pie charts, and nodes that had a low level of importance are represented by small pie charts. Specifically, the search items *vape pens*, *ENDS*, *e-cig*, *electronic cigarette*, *vaping*, *vape pods*, and *e-cigarette* had the highest BC (Table 1), whereas *disposable vape*, *box mods*, and *cigalikes* had the lowest values and, therefore, were the least important nodes in the network of search items.

Rule-based classification identified that *e-cigarette device* was dominant in all search items (34%-57%), followed by videos on e-liquids (14%-37%), cannabis (16%-28%), and *other* (8%-22%). These classes (labels) are shown in Figure 4. The pie chart, which represents a search item in the network of videos, shows the percentages of videos that belong to 1 of the 4 classes. For example, in the *disposable e-cigs* node, 57% of the videos are labeled as *e-cigarette device*, which is the highest among all the nodes (see Multimedia Appendix 1 for the percentage of classes for each search item). Interestingly, although we did not use search items that were directly related to *cannabis*, all the search items did have a considerable percentage of videos with vaping cannabis content.

Figure 5 shows the correlogram (correlation matrix), with each value indicating the correlation coefficient between the 2 corresponding search items. The magnitude of the correlation is represented by the color and a number on the corresponding square. Most pairs had negative or low positive correlations. The highest positive correlation in the network was observed between *electronic cigarette* and *ENDS* (0.75). The lowest correlation was observed between *vape mods* and *disposable e-cigs* (−0.28), which was the lowest in the network, suggesting that these nodes behave differently.

Figure 5. The correlogram of the search item network. Each cell indicates the correlation between the corresponding 2 nodes. The strength of the correlation is represented both by numerical value and by the gradient (represented by background color). More significant correlations with P≤.05 are shown in bold font. ENDS: electronic nicotine delivery system.



Discussion

Principal Findings

In this study, we used novel methods—such as network analysis and rule-based classification—to evaluate the associations among e-cigarette–related search items on YouTube. We observed that broad search items such as *e-cig* had the most connections to other search items and that specific search items such as *cigalike* had the least connections. Search items with similar words (eg, *vape* and *vaping*) and search items with similar meaning (eg, *e-liquid* and *e-juice*) yielded a high degree of connectedness. We also found that each node (ie, search item) had 18 (SD 34.8) connections (common videos) on average. BC indicated that general search items such as *electronic cigarette* and *vaping* had high importance in the network (0.00836). Our rule-based classification sorted videos into 4 categories: e-cigarette devices, cannabis vaping, e-liquid, and *other*.

Our results show that similar search items that were more specific (eg, *e-liquids* and *e-juice*) derived similar videos and that broad items (eg, *e-cig* and *e-cigarettes*) yielded a wide range of videos that were also identified with other search items. More specific items such as *cigalikes* and *pod mods* were less likely to be connected to other types of videos and identified unique videos that closely represented these items. For example, Massey et al [12] examined specific topics related to e-cigarettes, such as modifications of e-cigarettes, and searched YouTube for 28 phrases related to e-cigarette use as well as items related to modifications such as custom build, modification, and dripping. These research findings suggest that when identifying a specific topic area related to e-cigarettes, both broad items and specific items relevant to the topic area should be used. Our findings also suggest that there is a high level of redundancy in search results between pairs of similar search items such as *ENDS* and *e-cigarettes*, *vape* and *vaping*, and *vape juice* and *e-juice*, which suggests that redundant search items can be removed and the e-cigarette–related YouTube videos retrieved would largely be unaffected.

Our findings suggest that the local connectivity of mutual nodes in the network is important, and we looked at the relationships between a node and its neighbors regardless of its relationship

with the corpus. However, we were also interested in finding the search items whose removal from the network would most disrupt the network. We used BC to demonstrate the flow of information between different search items in the network of e-cigarette–related videos to accomplish this. In our information network (Figure 4), search items (nodes) were connected through common videos and, when there was no direct connection—illustrated by a line (edge)—between 2 nodes, a third node might enable connectedness. For instance, a search for *pod mods* on YouTube may return search results of videos that may also show search items such as *vape pods* and *vape mods*. When a user watches a video (retrieved from searching *pod mods*) that occurs in both *pod mods* and *vape pods*, the content of the video can influence the user and help the user consume *vape pods* content. However, as there are no direct common videos between *pod mods* and *disposable vape*, there is no content to guide the user directly toward videos related to *disposable vape* (from retrieved videos from searching *pod mods* only). Thus, a third search item (eg, *vape pods*) is needed for a user to traverse between these 2 nodes. Our network was relatively small but, for large networks, there might be several paths that users can traverse to reach nodes from a starting point in the network of search items. It should be noted that users can access information related to *disposable vape* through a variety of other related search items, such as *disposable*, *vape pens*, and *e-cig* (Figure 3). For the data that we collected, YouTube search results presented no common videos between the *pod mods* and *disposable vape* search items in the 16 fictitious profiles that we created. Ultimately, the search results were time dependent, and our results may not hold true in the future if YouTube's algorithms or input data change.

In a network, nodes that possess more central positions (higher BC) are more likely to provide informational content with less central members and have a more heterogeneous connection array [29,30]. In our study, search items such as *electronic cigarettes*, *ENDS*, and *vape pens* played this role. Looking at the measure of all search items, there was considerable variation in this value (from 0.00049 to 0.00836). However, with that noted, the overall BC values were low, meaning that connections in the network of search items could be made without intermediaries [31]. Generally, BC highlights the importance of a node in an information network as a transfer point between any pair of nodes. Consider the case in which a YouTube user uses several search items (nodes in our network) to obtain e-cigarette–related content (similar to the framework that we developed in this study). Some search items are more central in the list of search items as they will generate results broadly related to e-cigarettes, whereas more specific, niche items may be needed to access content specific to particular subdomains related to e-cigarettes.

Limitations and Strengths

The limitations of this study should be noted. Although we examined 18 search items related to e-cigarettes based on the items previously used in other studies [12], we did not use all the potential items used to refer to e-cigarettes to assess what video content users were likely to encounter on YouTube. There are many other items used to refer to e-cigarettes, such as brand-specific names (eg, *Juuling*) that were not used in this study and vernacular that is used to refer to e-cigarette use (eg, *stick*). Future work can use our methods to evaluate other search items to assess whether these search items produce relevant e-cigarette–related videos of interest. Our study findings indicate that search items that have similar meanings or words are likely to have significant overlap in the items of the resulting videos. Therefore, studies would benefit from pilot or exploratory work to determine the best search query items before proceeding with full-scale data collection. It is important to note that actual users search social media platforms for more complex themes rather than single words, such as motivations for using e-cigarettes or concerns about health outcomes related to their use [32]. Future research should extend our methods to assess the search results derived from a combination of words with different themes.

Our rule-based classification of video transcripts provides further information about the network. However, rule-based classification has some limitations as it relies on domain-specific expert knowledge. Therefore, each class may include some features (indicators similar to those in Textbox 1) that have not been incorporated into the selected features of that label. Despite these limitations, this is the first study to examine e-cigarette–related content on YouTube using NLP, video classification, and network modeling.

Another limitation of this study is that we used English-language search items to query the YouTube API. Other languages can be used to search videos on YouTube. Future work should evaluate other search items to assess whether our findings hold true for non–English-language videos.

Our study contributes to the existing literature related to e-cigarettes on YouTube [6,10,12,33,34] and provides tools to explore the search items as an information network. We developed a general framework for acquiring, cleaning, and analyzing the search data related to e-cigarettes on YouTube. Our use of network-based approaches presents a novel approach to the study of e-cigarettes on social media platforms. Our network-based approach highlighted the connections between different search items and enabled us to identify the associations between various e-cigarette–related search items. We used network analytic methods—particularly BC, which provided quantitative measures to better understand e-cigarette–related content on YouTube—to understand the underlying relationships between search items and characterize the information network formed by thousands of videos. Several centrality measures have been developed and are currently being used by search engines and recommendation systems in social media [35]. Although measures such as BC have been used extensively to study health-related content on social media [36], our study is the first to use BC to demonstrate the flow of information between different search items in the network of e-cigarette–related videos.

Importantly, our methods can be used by public health researchers to optimize search results on YouTube to better understand e-cigarette trends. Understanding search items is important as search items drive the information being presented to users. Our methods provide insight into how YouTube's algorithm selects and presents e-cigarette videos to users.

Surveillance of e-cigarettes on YouTube is crucial to understanding how health information and marketing are being communicated to users and nonusers, particularly among youth. Youth are vulnerable to e-cigarette content on social media [37], and they actively use social media platforms such as YouTube to obtain information, including information on novel uses of e-cigarettes [38]. Thus, having a better understanding of e-cigarette information being shown to youth is an important public health goal, as such information can provide insight into the public health policies needed and the social media platforms that control these algorithms. Given that YouTube's algorithm is proprietary, we can use these methods to evaluate search results even if YouTube introduces wholly new algorithms to evaluate the connectedness between search items and resulting videos. Our approach can also be applied longitudinally to determine whether YouTube search results change over time and to what extent the network of search items and their properties differ as a function of time. In addition, although we focused on e-cigarette–related content on YouTube in this study, we believe that our approach can be implemented on other platforms. Similarly, Harris et al [39] used network analysis to study the content of tweets and identify Twitter users prominent in the conversation for and against the e-cigarette Twitter campaign of the Chicago Department of Public Health. Future work could also evaluate whether our findings hold true for other social media platforms.

Conclusions

We found that similar search items (eg, *e-liquids* and *e-juice*) and items with similar word structures (eg, *e-cig* and *e-cigarettes*) yielded similar videos. In addition, general search items such as *e-cig* yielded a broad range of videos that were also identified by other related search items. Ultimately, broader search items act as gateways to the network of e-cigarette–related search items. More specific items such as *cigalikes* and *pod mods* are less likely to be connected to other types of videos and are useful in identifying unique videos that may closely represent more specific, niche e-cigarette–related areas. Our methods can be used as a measure to exclude some search items from future studies, restrict the number of search items, and identify search items that serve as important gateways to broader e-cigarette–related content. Importantly, public health researchers can use our methods to optimize search results on YouTube to better understand how search items related to e-cigarettes drive the content that is shown to youth on this popular social media platform as well as on other social media platforms.

Acknowledgments

This study was supported by the National Institutes of Health's National Institute on Drug Abuse (award number R01DA049878).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Number of videos retrieved from each search item and the percentage of each class. We retrieved 5875 videos (4201 unique) by searching for 18 search items. There were 1674 duplicate videos between search items.
[DOCX File , 15 KB - jmir_v24i1e30679_app1.docx]

References

1. Cornelius ME, Wang TW, Jamal A, Loretan CG, Neff LJ. Tobacco product use among adults - United States, 2019. MMWR Morb Mortal Wkly Rep 2020 Nov 20;69(46):1736-1742 [FREE Full text] [doi: 10.15585/mmwr.mm6946a4] [Medline: 33211681]
2. Wang TW, Neff LJ, Park-Lee E, Ren C, Cullen KA, King BA. E-cigarette use among middle and high school students - United States, 2020. MMWR Morb Mortal Wkly Rep 2020 Sep 18;69(37):1310-1312 [FREE Full text] [doi: 10.15585/mmwr.mm6937e1] [Medline: 32941408]
3. Kapan A, Stefanac S, Sandner I, Haider S, Grabovac I, Dorner T. Use of electronic cigarettes in European populations: a narrative review. Int J Environ Res Public Health 2020 Mar 17;17(6):1971 [FREE Full text] [doi: 10.3390/ijerph17061971] [Medline: 32192139]
4. Wang X, Zhang X, Xu X, Gao Y. Perceptions and use of electronic cigarettes among young adults in China. Tob Induc Dis 2019 Mar 4;17(March):17 [FREE Full text] [doi: 10.18332/tid/102788] [Medline: 31582928]
5. Auxier B, Anderson M. Social media use in 2021. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/2021/04/07/social-media-use-in-2021/ [accessed 2021-09-21]
6. Huang J, Kornfield R, Emery SL. 100 million views of electronic cigarette YouTube videos and counting: quantification, content evaluation, and engagement levels of videos. J Med Internet Res 2016 Mar 18;18(3):e67 [FREE Full text] [doi: 10.2196/jmir.4265] [Medline: 26993213]
7. Statistics. YouTube. URL: https://www.youtube.com/yt/about/press/ [accessed 2021-05-06]
8. Paek H, Kim S, Hove T, Huh JY. Reduced harm or another gateway to smoking? Source, message, and information characteristics of E-cigarette videos on YouTube. J Health Commun 2014 Oct;19(5):545-560. [doi: 10.1080/10810730.2013.821560] [Medline: 24117370]

9. Sears CG, Walker KL, Hart JL, Lee AS, Siu A, Smith C. Clean, cheap, convenient: promotion of electronic cigarettes on YouTube. Tob Prev Cessation 2017 Apr 07;3(10):69393 [FREE Full text] [doi: 10.18332/tpc/69393]
10. Kong G, LaVallee H, Rams A, Ramamurthi D, Krishnan-Sarin S. Promotion of vape tricks on YouTube: content analysis. J Med Internet Res 2019 Jun 18;21(6):e12709 [FREE Full text] [doi: 10.2196/12709] [Medline: 31215510]
11. Guy M, Helt J, Palafox S, Green K, Soule E, Maloney S, et al. Orthodox and unorthodox uses of electronic cigarettes: a surveillance of YouTube video content. Nicotine Tob Res 2019 Sep 19;21(10):1378-1384 [FREE Full text] [doi: 10.1093/ntr/nty132] [Medline: 29961828]
12. Massey ZB, Li Y, Holli J, Churchill V, Yang B, Henderson K, et al. Modifications to electronic nicotine delivery systems: content analysis of YouTube videos. J Med Internet Res 2020 Jun 02;22(6):e17104 [FREE Full text] [doi: 10.2196/17104] [Medline: 32348288]
13. Hua M, Yip H, Talbot P. Mining data on usage of electronic nicotine delivery systems (ENDS) from YouTube videos. Tob Control 2013 Mar;22(2):103-106. [doi: 10.1136/tobaccocontrol-2011-050226] [Medline: 22116832]
14. Luo C, Zheng X, Zeng DD, Leischow S. Portrayal of electronic cigarettes on YouTube. BMC Public Health 2014 Oct 03;14:1028 [FREE Full text] [doi: 10.1186/1471-2458-14-1028] [Medline: 25277872]
15. Jones DM, Guy MC, Soule E, Sakuma KK, Pokhrel P, Orloff M, et al. Characterization of electronic cigarette warning statements portrayed in YouTube videos. Nic Tob Research 2021 Jan 05;23(8):1358-1366. [doi: 10.1093/ntr/ntaa272]
16. Merianos AL, Gittens OE, Mahabee-Gittens EM. Depiction of health effects of electronic cigarettes on YouTube. J Subst Use 2016 May 09;21(6):614-619 [FREE Full text] [doi: 10.3109/14659891.2015.1118565] [Medline: 28217030]
17. Paolillo J. Structure and network in the YouTube core. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008). 2008 Presented at: 41st Annual Hawaii International Conference on System Sciences (HICSS 2008); Jan. 7-10, 2008; Waikoloa, HI, USA p. 156. [doi: 10.1109/hicss.2008.415]
18. Abul-Fottouh D, Song MY, Gruzd A. Corrigendum to "Examining algorithmic biases in YouTube's recommendations of vaccine videos" [Int. J. Med. Inf. 140 (2020) 104175]. Int J Med Inform 2021 Apr;148:104385 [FREE Full text] [doi: 10.1016/j.ijmedinf.2021.104385] [Medline: 33515931]
19. Murthy D. Evaluating platform accountability: terrorist content on YouTube. Am Behav Scient 2021 Feb 03;65(6):800-824. [doi: 10.1177/0002764221989774]
20. Bockholt M, Zweig K. A systematic evaluation of assumptions in centrality measures by empirical flow data. Soc Netw Anal Min 2021 Mar 01;11(1):1-30 [FREE Full text] [doi: 10.1007/s13278-021-00725-3]
21. Harlow A, Stokes A, Brooks D. Socioeconomic and racial/ethnic differences in e-cigarette uptake among cigarette smokers: longitudinal analysis of the Population Assessment of Tobacco and Health (PATH) Study. Nicotine Tob Res 2019 Sep 19;21(10):1385-1393 [FREE Full text] [doi: 10.1093/ntr/nty141] [Medline: 29986109]
22. Orbot: Tor for Android. Google Play Store. URL: https://play.google.com/store/apps/details?id=org.torproject.android&hl=en_US&gl=US [accessed 2021-09-21]
23. Behera RK, Naik D, Ramesh D, Rath SK. MR-IBC: MapReduce-based incremental betweenness centrality in large-scale complex networks. Soc Netw Anal Min 2020 Apr 12;10(1):1-3. [doi: 10.1007/s13278-020-00636-9]
24. Yan C, Zhang X, Gao C, Wilfong E, Casey J, France D, et al. Collaboration structures in COVID-19 critical care: retrospective network analysis study. JMIR Hum Factors 2021 Mar 08;8(1):e25724 [FREE Full text] [doi: 10.2196/25724] [Medline: 33621187]
25. spaCy: Industrial-strength Natural Language Processing in Python. spaCy. URL: https://spacy.io/ [accessed 2021-05-06]
26. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); October, 2014; Doha, Qatar p. 1532-1543 URL: https://nlp.stanford.edu/projects/glove/ [doi: 10.3115/v1/d14-1162]
27. Rivas R, Sadah SA, Guo Y, Hristidis V. Classification of health-related social media posts: evaluation of post content-classifier models and analysis of user demographics. JMIR Public Health Surveill 2020 Apr 01;6(2):e14952. [doi: 10.2196/14952] [Medline: 32234706]
28. Sheikhalishahi S, Miotto R, Dudley JT, Lavelli A, Rinaldi F, Osmani V. Natural language processing of clinical notes on chronic diseases: systematic review. JMIR Med Inform 2019 Apr 27;7(2):e12239 [FREE Full text] [doi: 10.2196/12239] [Medline: 31066697]
29. Wang H, Hong M. Supervised Hebb rule based feature selection for text classification. Inform Process Manag 2019 Jan;56(1):167-191. [doi: 10.1016/j.ipm.2018.09.004]
30. Pan W, Feng B, Shen C. Examining social capital, social support, and language use in an online depression forum: social network and content analysis. J Med Internet Res 2020 Jun 24;22(6):e17365 [FREE Full text] [doi: 10.2196/17365] [Medline: 32579125]
31. Kaonga NN, Labrique A, Mechael P, Akosah E, Ohemeng-Dapaah S, Baah JS, et al. Using social networking to understand social networks: analysis of a mobile phone closed user group used by a Ghanaian health team. J Med Internet Res 2013 Apr 03;15(4):e74 [FREE Full text] [doi: 10.2196/jmir.2332] [Medline: 23552721]
32. Kwon M, Park E. Perceptions and sentiments about electronic cigarettes on social media platforms: systematic review. JMIR Public Health Surveill 2020 Jan 15;6(1):e13673 [FREE Full text] [doi: 10.2196/13673] [Medline: 31939747]

33. Albarracin D, Romer D, Jones C, Hall Jamieson K, Jamieson P. Misleading claims about tobacco products in YoutTube videos: experimental effects of misinformation on unhealthy attitudes. J Med Internet Res 2018 Jun 29;20(6):e229 [FREE Full text] [doi: 10.2196/jmir.9959] [Medline: 29959113]
34. Cho H, Li W, Shen L, Cannon J. Mechanisms of social media effects on attitudes toward e-cigarette use: motivations, mediators, and moderators in a national survey of adolescents. J Med Internet Res 2019 Jun 27;21(6):e14303 [FREE Full text] [doi: 10.2196/14303] [Medline: 31250830]
35. Wang L, Wang J, Wang M, Li Y, Liang Y, Xu D. Using internet search engines to obtain medical information: a comparative study. J Med Internet Res 2012;14(3):e74 [FREE Full text] [doi: 10.2196/jmir.1943] [Medline: 22672889]
36. Meza XV, Yamanaka T. Food communication and its related sentiment in local and organic food videos on YouTube. J Med Internet Res 2020 Aug 10;22(8):e16761 [FREE Full text] [doi: 10.2196/16761] [Medline: 32773370]
37. Zheng X, Li W, Wong S, Lin H. Social media and e-cigarette use among US youth: longitudinal evidence on the role of online advertisement exposure and risk perception. Addict Behav 2021 Aug;119:106916. [doi: 10.1016/j.addbeh.2021.106916] [Medline: 33798917]
38. Kong G, Morean ME, Bold KW, Wu R, Bhatti H, Simon P, et al. Dripping and vape tricks: alternative e-cigarette use behaviors among adolescents. Addict Behav 2020 Aug;107:106394 [FREE Full text] [doi: 10.1016/j.addbeh.2020.106394] [Medline: 32222561]
39. Harris JK, Moreland-Russell S, Choucair B, Mansour R, Staub M, Simmons K. Tweeting for and against public health policy: response to the Chicago Department of Public Health's electronic cigarette Twitter campaign. J Med Internet Res 2014;16(10):e238 [FREE Full text] [doi: 10.2196/jmir.3622] [Medline: 25320863]

Abbreviations

API: application programming interface
BC: betweenness centrality
NLP: natural language processing

Edited by R Kukafka; submitted 24.05.21; peer-reviewed by A Bailey, T Ndabu, MS Hosseini, V Moquillaza Alcántara; comments to author 30.07.21; revised version received 27.09.21; accepted 29.10.21; published 27.01.22.

Please cite as:
Dashtian H, Murthy D, Kong G
An Exploration of e-Cigarette–Related Search Items on YouTube: Network Analysis
J Med Internet Res 2022;24(1):e30679
URL: https://www.jmir.org/2022/1/e30679
doi:10.2196/30679
PMID:35084353

Original Paper

Using the COVID-19 Pandemic to Assess the Influence of News Affect on Online Mental Health-Related Search Behavior Across the United States: Integrated Sentiment Analysis and the Circumplex Model of Affect

Damien Lekkas[1,2], MS; Joseph A Gyorda[1]; George D Price[1,2], BS; Zoe Wortzman[1]; Nicholas C Jacobson[1,2,3,4], PhD

[1]Center for Technology and Behavioral Health, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States

[2]Quantitative Biomedical Sciences Program, Dartmouth College, Hanover, NH, United States

[3]Department of Biomedical Data Science, Geisel School of Medicine, Dartmouth College, Lebanon, NH, United States

[4]Department of Psychiatry, Geisel School of Medicine, Dartmouth College, Hanover, NH, United States

Corresponding Author:
Damien Lekkas, MS
Center for Technology and Behavioral Health
Geisel School of Medicine
Dartmouth College
46 Centerra Parkway
Suite 300, Office #313S
Lebanon, NH, 03766
United States
Phone: 1 603 646 7000
Email: Damien.Lekkas.GR@dartmouth.edu

Abstract

Background: The digital era has ushered in an unprecedented volume of readily accessible information, including news coverage of current events. Research has shown that the sentiment of news articles can evoke emotional responses from readers on a daily basis with specific evidence for increased anxiety and depression in response to coverage of the recent COVID-19 pandemic. Given the primacy and relevance of such information exposure, its daily impact on the mental health of the general population within this modality warrants further nuanced investigation.

Objective: Using the COVID-19 pandemic as a subject-specific example, this work aimed to profile and examine associations between the dynamics of semantic affect in online local news headlines and same-day online mental health term search behavior over time across the United States.

Methods: Using COVID-19–related news headlines from a database of online news stories in conjunction with mental health–related online search data from Google Trends, this paper first explored the statistical and qualitative affective properties of state-specific COVID-19 news coverage across the United States from January 23, 2020, to October 22, 2020. The resultant operationalizations and findings from the joint application of dictionary-based sentiment analysis and the circumplex theory of affect informed the construction of subsequent hypothesis-driven mixed effects models. Daily state-specific counts of mental health search queries were regressed on circumplex-derived features of semantic affect, time, and state (as a random effect) to model the associations between the dynamics of news affect and search behavior throughout the pandemic. Search terms were also grouped into depression symptoms, anxiety symptoms, and nonspecific depression and anxiety symptoms to model the broad impact of news coverage on mental health.

Results: Exploratory efforts revealed patterns in day-to-day news headline affect variation across the first 9 months of the pandemic. In addition, circumplex mapping of the most frequently used words in state-specific headlines uncovered time-agnostic similarities and differences across the United States, including the ubiquitous use of negatively valenced and strongly arousing language. Subsequent mixed effects modeling implicated increased consistency in affective tone ($Spin_{VA}$ β=–.207; P<.001) as predictive of increased depression-related search term activity, with emotional language patterns indicative of affective uncontrollability ($Flux_{A}$ β=.221; P<.001) contributing generally to an increase in online mental health search term frequency.

Conclusions: This study demonstrated promise in applying the circumplex model of affect to written content and provided a practical example for how circumplex theory can be integrated with sentiment analysis techniques to interrogate mental health–related associations. The findings from pandemic-specific news headlines highlighted arousal, flux, and spin as potentially significant affect-based foci for further study. Future efforts may also benefit from more expansive sentiment analysis approaches to more broadly test the practical application and theoretical capabilities of the circumplex model of affect on text-based data.

(J Med Internet Res 2022;24(1):e32731) doi:10.2196/32731

KEYWORDS

affect; sentiment; circumplex; news; mental health; online search behavior; generalized mixed models; natural language processing; anxiety; depression; coronavirus; internet; information seeking; behavior; online health information; COVID-19

Introduction

News coverage can have a significant impact on mental health. In particular, the sentiment of new articles, or their views and attitudes toward specific topics, can evoke emotional responses in consumers [1-7]. For instance, news articles with a negative sentiment can elicit negative emotions on a daily basis [1], and, similarly, news articles with a positive sentiment can elicit positive emotions and enjoyment in consumers [2]. The intensity of the emotional response to the news is strongly related to the personal relevance of stories [1,3]. Emotional intensity is also contingent on the type of news: forms of "hard" news such as significant world events and more pressing issues (eg, natural disasters, political turmoil) typically evoke stronger responses than "soft" news articles that are more sensational and less timely (eg, sports, pop culture, tabloid-style stories) [2,4]. Furthermore, online sources such as social media are being increasingly used as a source of news [8], and posts seen in news feeds have shown to elicit emotional responses aligning with the sentiment of the posts [5]. Given the impact of daily emotional responses on long-term mental health (eg, [9]), it is worth investigating the role of daily stressors such as news coverage sentiment in the mental health of consumers.

News headlines are a specific facet of news coverage that can impact mental health. Headlines attempt to draw attention to specific details and facts about the main story, often significantly impacting how the reader perceives the given information [6]. To attract the reader, news headlines will often contain strong sentiment [6,10], leading readers to experience emotional responses from simply reading headlines [7]. Furthermore, social media outlets such as Facebook are being increasingly used as news sources, and estimates suggest that over 90% of Facebook users only read the headlines of news stories in their feeds [11]. Thus, while both story headlines and content are important to consider, specific focus on headlines may be more prescient from a mental health perspective.

A recent global issue of pervasive personal relevance and therefore particular noteworthiness when considering the mental health implications of news coverage is the COVID-19 pandemic. Recent studies have reported that people respond adversely to news pertaining to COVID-19. One study found that over one-third of participants were spending at least two hours on social media reading COVID-19–related news, and extended exposure to COVID-19 news was associated with higher anxiety and depression in adults [12]. More extreme exposure (eg, ≥3 hours) has been found to be associated with generalized anxiety disorder [13]. Studies have led to similar findings among college-aged students, reporting that students spending an hour or more online looking for COVID-19–related news have significantly higher levels of anxiety and somatization [14]. In a separate multimodal passive sensing study with self-reported ecological momentary assessments conducted on 217 undergraduate students in the United States, students were found to be more sedentary and reported increased anxiety and depression symptoms (P<.001) during the academic term coinciding with the first few months of the COVID-19 pandemic (Winter 2020) compared with previous academic terms [15]. Additionally, movement and sleep-related behaviors were found to be associated with fluctuations in COVID-19 reporting [15]. Furthermore, another study looked specifically at the sentiment of headlines in COVID-19 news articles and found that over 50% of headlines had a negative sentiment, 30% had a positive sentiment, and under 20% had a neutral or nonpolarizing sentiment [16]. Taken together, there is empirical precedent to link the consumption of COVID-19–related news to changes in mental health.

Given that the majority of people under 25 years perform internet searches as their primary method of seeking mental health information and help [17], it stands to reason that internet search behavior may be a novel and powerful indicator of mental health changes. With readily available data accessible via platforms such as Google Trends, internet search behavior represents a large corpus of data with increasing prevalence in epidemiological applications. For example, peaks in suicide search term activity correspond to high-profile news stories pertaining to suicide [18] as well as completed suicide rates in many countries [19]. Along with this, internet searches exhibit seasonal variability, with peaks in mental health searches occurring in the winter and troughs occurring in the summer, aligning with seasonal depressive disorder [20]. Furthermore, mental health searches in the United States spiked in early 2020 after lockdown announcements due to the COVID-19 pandemic, but such searches leveled off in response to the issuing of stay-at-home orders [21]. All in all, the leverage of records derived from internet search behavior represents a broader method of sampling that may have implications for population-level mental health—a degree of heterogeneity that is not always attainable through more traditional epidemiological or clinical sampling approaches.

Given the versatility of internet search trends as an indicator of mental health changes, a popular and successful approach to analyzing such text corpora is using natural language processing

(NLP). NLP is a computational-based approach to analyzing text, and it has been used to study and model a variety of mental health constructs [22-27]. A particularly useful facet of NLP is sentiment analysis, which is a method of computationally assessing opinions, subjectivity, and emotion in text [22]. At a basic level, sentiment analysis constitutes the extraction of words and phrases (n-grams) from a text corpus to compare these features with prior knowledge regarding their affective connotation. A common approach to extraction is to use a rule-based technique, which applies a consistent system of manually derived rules for reference and identification of sentiment, such as a dictionary mapping of words to a specific sentiment category or value (eg, positive or negative) [23]. Studies leveraging sentiment analysis have also relied on automatic techniques that operate by implementing machine learning–based methods, which classify the sentiment of novel text by "learning" from the features of example data (eg, [24]). In more recent literature, hybrid techniques combining elements of both rule-based and automatic techniques have emerged, often being used for analyzing sentiment across domains (eg, [25,26]). Sentiment analysis has been commonly applied to text from social media platforms such as Twitter. For instance, Chakraborty et al [27] analyzed the sentiment of tweets to assess mental health in response to COVID-19, and Chintalapudi et al [28] applied a hybrid-based sentiment analysis using a naïve Bayes classifier to patient medical records to assess trends in patients' physical and mental health over time. In the broadest sense, this methodology has presented as a promising means to quantitatively ascertain mental health dynamics.

The circumplex model of affect is a theory-driven approach to ascertain emotional dynamics from the quantitative perspective of sentiment analysis. One version of the model posits that, for each individual, all affective states stem from 3 neurophysiological states: valence (pleasure/displeasure), arousal (alertness/excitation), and dominance (autonomy/restriction) [29]. Specifically, all emotions can be represented as a linear combination of valence, arousal, and dominance, or different degrees of these states [29]. It is worth noting that affect, sentiment, and emotion, while they are often regarded as equivalent and used interchangeably, are distinct in the psychological literature. Affect is considered an "umbrella term" that encompasses both emotion and sentiment which refers to an individual's subjective response to an experience, which can vary in positive and negative intensity [30]. Emotion is a conscious affective experience that is an adaptive response to some event which, according to circumplex theory, can be measured by its valence and arousal [30]. Finally, sentiment is an individual's acquired affective disposition toward an object or event, wherein a disposition will have emotions associated with it [30]. Thus, the affect circumplex inherently refers to both sentiment and emotion. Preceding the valence–arousal–dominance (VAD-3D) circumplex model is the valence–arousal (VA-2D) circumplex model [31]. While the VA-2D model has the earliest and widest empirical support in the literature (eg, [31,32]), the VAD-3D model has shown to be more sensitive in the differentiation of certain affective states (eg, fear and anger) compared with the VA-2D model (eg, [29,33]). In general, a circumplex model can be constructed for an individual by assessing affective states after multiple interactions over time. Further analysis of the circumplex dimensions offer deeper understanding of an individual's emotions and emotion dynamics, or how an individual's emotions change over time.

Although the affect circumplex has traditionally been used to temporally model emotion in individuals, it has also been used to operationalize word usage within the context of sentiment analysis. For instance, the circumplex model has been applied to detect sentiment from social media content, including Twitter tweets [34] and Facebook posts [35]. While there have been several disparate and effective empirical applications of the affect circumplex to the greater sentiment analysis framework, the properties of the circumplex itself have also been utilized within the research domain of emotion dynamics. In a meta-analysis, researchers examined the flux, pulse, and spin of affect [36] to predict changes in aggressive behaviors and found that flux in positive affect, contrary to expectation, was associated with individual aggression [37]. More broadly, their findings supported the application of such dynamic measures to better understand human social behavior [37]. While currently underrepresented in the literature as a suite of metrics within the NLP toolkit, there is broader empirical precedent to apply the affect circumplex as a uniquely synergistic operationalization of emotion/sentiment. Inspired by the promise of circumplex theory and application within the extant literature, this study aimed to integrate notions of circumplex dynamics with a medium of the written word, thereby leveraging novel quantifications of semantics for exploratory and hypothesis-testing applications within the mental health space.

The COVID-19 pandemic, as of June 2021, has claimed the lives of nearly 4 million individuals worldwide [38], significantly altering the infrastructure, enterprise, and sociobehavioral landscape of the globe. Its widespread, devastating impact to physical, mental, and economic well-being has presented a unique case study opportunity to further explore how news media coverage of such a ubiquitous, baleful issue influences the mental health of a society that is now highly integrated with, and reliant on, digital technology for information exchange. The breadth and depth of the data associated with analyses on this front necessitate creative and novel analytical approaches to uncover nuanced patterns that may have potential implications for how information is expressed as well as for the well-being of those who are exposed to its expression. Given the broadly unifying and consistently impactful nature of the recent pandemic, the current study sought to interrogate the association between the affective dynamics in digitally accessible local news media and online trends in mental health search behavior across the United States. This work capitalized on the affordances of online "big data" and leveraged both Google Trends and Media Cloud repositories to collect daily, state-specific, concurrent information on internet search activity and written news media coverage of the pandemic, respectively. Data from March 24, 2020, to October 22, 2020, were used to specifically profile COVID-19–related news article headlines based on affect. As mentioned in previous work, a focus on news headlines provides several benefits: (1) they are dense with contextual information while being short and more consistently formatted, (2) they often appeal to

readers' emotions, and (3) avoid issues of data privacy [7]. Relying on NLP techniques to ultimately draw from the circumplex theory of affect and associated operationalizations of emotion dynamics, in conjunction with generalized mixed statistical modeling to assess online mental health behavioral correlations with these operationalizations, this research contributes a novel, integrative analytical framework for the exploration and quantification of news media affect in the digital era.

To this end, the current body of work consisted of a descriptive/exploratory aim leveraged to inform subsequent research questions aimed to model online mental health–based search behavior in a targeted and empirically justified manner. Accordingly, the first aim was to gain some intuition regarding the language surrounding written COVID-19 news reporting and characterize the semantic affect of written COVID-19 news reporting across the United States. This aim was expressed in terms of 2 initial exploratory research questions:

1. How does affect (valence, arousal, and dominance) of state-specific COVID-19–related news headlines change throughout the course of the initial phases of the pandemic?
2. How different is overall word usage in COVID-19–related news headlines across the country?

The statistical and qualitative answers to these questions at baseline were then used to inform and justify subsequent modeling decisions for the second aim, namely, to interrogate the association between mental health–based online search behavior and the affective dynamics of COVID-19 news exposure through time. Specifically, the resulting models leveraged operationalized word affect exposure and focused on the potential utility of affective dynamics in circumplex space to model changes in online mental health search behavior through time. As a result, this second aim was driven by the following 2 additional research questions:

1. Do variables that reflect dimensional circumplex dynamics of affect (flux, pulse, and spin) across daily news story headlines have a stronger, statistically significant association with same-day online mental health search behavior outcomes compared with more simple measures of affect (ie, daily average valence, arousal, and dominance)?
2. Given the core importance of valence and arousal in affect theory, are features that capture the valence and arousal components of news headlines more significant and more highly associated with mental health search behavior outcomes compared with dominance-related metrics?

Methods

Study Sample

This work used publicly available COVID-19–related online news headlines. As described in more detail in the "News Data" section, 88,987 news story titles were analyzed in total and represented written coverage of the pandemic through time across the top-ranked state-based news outlets by web traffic in the United States (n=135 news outlets).

Data Collection

News Data

News outlets were selected to represent daily sources that have the largest audience while also emphasizing coverage of local state-wide news, thus broader news outlets with substantial national or international coverage were not selected (eg, the New York Times, the Washington Post, USA Today). Web traffic was used as a heuristic to select representative news outlets on a per-state basis, with the requirement that the source ranked within the top 5 of all outlets for that state [39]. Sources were not considered if they were not among the most frequently visited. In many cases, sources that ranked highly in terms of web traffic were also ranked within the top 10 in terms of physical circulation [40] (Multimedia Appendix 1). As the study aimed to investigate the association between online search behavior (operationalized through Google Trends data) and the language of news coverage, it was most practical to focus on news information readily available via the web. To this end, the study utilized the Media Cloud API client, an open-source platform for tracking millions of stories published online, to ultimately construct a web scraping pipeline in Python and programmatically extract the titles, dates, and body text from stories of selected news sources [41,42]. COVID-19–related news stories were queried from January 23, 2020, to October 22, 2020, for the presence of "covid", "covid19", "covid-19", "coronavirus", or "pandemic" in at least one of either the story headline or main body text.

The study aimed for 3 news sources per state, but due to the limitations of data availability via Media Cloud as well as paywalls specific to certain news outlets, several states did not have appreciable representation of stories from top-ranked outlets available for access between the dates specified for the study. Typically, if Media Cloud had insufficient data for a news source (defined as <15 stories per month on average), the news source with the next highest web traffic rate was chosen (up to the fifth highest ranked news source). An exhaustive list of the sources used and their respective Media Cloud identifiers is provided in Multimedia Appendix 1.

The analyses for this work were conducted specifically on the content of story headlines. All stories were manually curated to remove advertisements and other non-news content as well as duplicate entries within any 1 news source. Stories without an associated date of publication were removed. In several cases, body text was not accessible; however, the headlines of these stories were still included for analysis. The numbers reported in Multimedia Appendix 1 represent the outlet-specific totals for analysis after curation and preprocessing. Despite the availability of news data from January 23, 2020, only data starting from March 24, 2020, to October 22, 2020, were used for downstream hypothesis testing, as the initiation of data collection associated with the outcomes of interest (see the "Mental Health Search Activity" section) occurred on March 24, 2020.

Mental Health Search Activity

Counts of mental health–related search terms were collected daily using the *gtrendsR* (v1.4.8) package in the R programming

language. For each of 17 mental health terms ("anxiety," "depression," "hopeless," "angry," "afraid," "apathy," "worthless," "worried," "restless," "irritable," "tense," "scattered," "tired," "avoiding," "insomnia," "suicidal", and "suicide"), queries were conducted for search activity across the United States with state-level resolution. Following the authors' previous work in this domain [21], these terms were adapted and validated in part by prior research on mental health using Google Trends [20] as well as by research that assessed rapid affective symptom changes in accordance with the Diagnostic and Statistical Manual of Mental Disorders, 5th edition [43,44].

As discussed in more detail in the Google Trends documentation, search data are automatically normalized and scaled from 0 to 100 based on a topic's proportion to all searches on all topics within a specified period and geographic location. However, for this work, raw counts of searches were desired rather than normalized values. To ultimately arrive at estimated raw counts of mental health term searches, queries in *gtrendsR* were constructed to include a unique daily comparator term that was selected to serve as a representative top-trending term for that day. These trending terms are published daily by Google in tiers of absolute frequency (eg, >50,000 hits) and are therefore not normalized. Each programmatic call to *gtrendsR* therefore consisted of 2 simultaneous keywords (eg, "anxiety" and "NASA") corresponding to a mental health term and the selected comparator term for that day. Manual selection of comparator terms was performed to ensure that the term was as unrelated as possible with regard to both COVID-19 and mental health. The complete list of utilized comparator terms by date, along with their estimated search volume thresholds, is provided in Multimedia Appendix 2. Ultimately, each query to *gtrendsR* returned state-level normalized counts for the target mental health term and the representative comparator term for the date of interest. From this information, a state-specific estimated daily count for each mental health term was calculated in the following manner:

$$MH_{NormTot} \times (COMP_{SV})/(COMP_{NormTot}) = MH_{AdjTot} \quad (1)$$

$$MH_{AdjTot} \times (MH_{NormState})/(MH_{NormTot}) \times (POP_{State})/(POP_{Tot}) = MH_{AdjState} \quad (2)$$

In Equation 1, an adjusted count for the target mental health term MH_{AdjTot} was first calculated by scaling the *gtrendsR* normalized total count of the target mental health term ($MH_{NormTot}$) by the ratio of the reported comparator term search volume to its *gtrendsR* normalized total count ($COMP_{SV}/COMP_{NormTot}$). Because Google reports search volume in tiers (eg, 50,000+, 100,000+), the value of $COMP_{SV}$ was selected to be the lowest possible value that represents that tier (eg, an estimated search volume of 50,000+ was set to correspond to a $COMP_{SV}$=50,000). MH_{AdjTot} is summative across all 50 states. To distill a practical estimation of mental health search term counts at the state level, Equation 2 then takes the ratio of the normalized target state count to the normalized total count (across all 50 states) for the target mental health term ($MH_{NormState}/MH_{NormTot}$) as well as the estimated ratio of the population belonging to the target state (POP_{State}/POP_{Tot}) (calculated from information provided by the 2019 US Census Bureau) and computes the product of these values to ultimately calculate the target state–specific portion of MH_{AdjTot} and yield $MH_{AdjState}$. Resulting counts were rounded down to the nearest whole number. The process was repeated for each of the 17 mental health terms, for each of the 50 states, across the 213 days of the study spanning from March 24, 2020, to October 22, 2020. Quantification and preprocessing of raw *gtrendsR* count data were performed with custom scripts written in the Python programming language (version 3.8.3).

Sentiment Analysis

Quantification of the emotional tone of words is a major analytical arm of NLP. A component and feature of many NLP analytical suites such as Python's *nltk* library [45], sentiment analysis commonly involves quantification of both polarity (valence) and intensity (arousal) of emotion. While it is common to employ pretrained models or dictionaries with packaged implementations such as what is provided through *nltk*, it is also possible to conduct sentiment analysis with a custom reference dictionary so as to capture additional or unique nuances of emotion. Accordingly, this work leveraged a large, comprehensive, and empirically derived reference dictionary of 13,915 English lemmas quantified by a crowdsourcing effort involving 1827 responders and over 1 million word ratings of valence, arousal, and dominance [46]. Average (SD) valence of all rated words in the dictionary is 5.064 (1.275), average arousal is 4.211 (0.986), and average dominance is 5.185 (0.938) on a scale of 1-9, with higher values indicative of more positive polarity, stronger intensity of emotion, or more dominant language. For this study, all individual word ratings of valence, arousal, and dominance were standardized to obtain Z-score equivalents prior to application in downstream analyses. As a result, numerical values represent SDs below or above the average (Z=0) across all reference dictionary words for that affective dimension.

With a custom-built Python script, tabular data containing a state's news story titles and dates of publication were first preprocessed using *nltk* and regular expressions to lowercase all words, remove all nonalphabetical characters, and remove common stop words—words that do not add meaning or provide context to a sentence (eg, "the," "she," "it"). Additional words were manually appended to the default stop word dictionary to account for location names (eg, "orange," "lake," "phoenix"), media language (eg, "video," "subscribe," "share"), and proper nouns doubling as reference dictionary entries (eg, "mitt," "trump"). A complete list of manually added stop words is available in Multimedia Appendix 3 for reference. News stories were then grouped by date of publication, and all words across story titles were pooled into date-based lists representing a given state's headline word usage across time. From these lists, the average Z-score valence, arousal, and dominance across words that mapped to the reference dictionary were calculated per day. This process was repeated for each of the 50 states to arrive at state-specific, daily averages of normalized news headline sentiment across 3 primary affect dimensions. Out of the total 731,607 non–stop words across all news story titles in the data

set, 252,524 (34.52%) mapped to the reference dictionary and were therefore considered in analyses.

Descriptive Statistics of State-Specific Affect Through Time

To address the exploratory component of this study, several approaches were taken. With regard to the first question of affect across time, the news collection period (January 23, 2020-October 22, 2020) was first divided into 5 distinct temporal windows comprising 3 phases of the pandemic in the United States. The first phase is referred to as "Pre-pandemic" and spans 50 days (January 23, 2020-March 12, 2020). The termination of this phase and the initiation of the second phase was defined by the official presidential announcement (Proclamation 9994) defining COVID-19 as a national emergency on March 13, 2020. The second phase, "Early Response," spans 26 days and terminates on April 7, 2020. This date marks South Carolina's announcement of stay-at-home orders, the last of 42 states to make an official statement regarding the required alterations to daily behavior and movement. The third and final phase, "Mid-pandemic," continues for 198 days to the end of the study period (April 8, 2020-October 22, 2020). This phase was further subdivided into 3 equal partitions of 66 days to create event-agnostic temporal parity *within* the phase. This also allowed for less biased statistical comparisons *across* all 3 phases because the resulting statistics summarized phases with more equivalent data representation.

News headlines were then parsed based on publication date into these 5 temporal bins. For each state, the (1) mean, (2) variance, and (3) root mean square of successive differences (RMSSDs) [47] for the standardized daily averaged valence, arousal, and dominance were calculated. Where mean and variance calculations represented more traditional summative statistical parameters and defined a phase more broadly, RMSSD captured notions of change and consistency in affect within temporal partitions over time. With the resulting data in hand (see Multimedia Appendix 4 for each state's partition-specific data), affect-specific choropleths for each of the 5 temporal windows were constructed. In addition, 6 representative states were selected (Arizona, Georgia, Illinois, New Jersey, Texas, and Washington) to further illustrate similarities and differences in state-specific affect mean and RMSSD across the designated phases of the pandemic. These states were chosen for further illustration based on their collective geopolitical breadth as well as their appreciable volume of available news headlines; however, all states were analyzed in subsequent modeling efforts (see the "Generalized Mixed Effects Modeling" section). The results of these initial collective efforts served in part as the impetus for further pattern-based operationalizations (see the "Affect Circumplex Feature Engineering and Exploratory Visualization " section) and hypothesis testing (see the "Generalized Mixed Effects Modeling" section).

Affect Circumplex Feature Engineering and Exploratory Visualization

Overview

Appreciable patterns and variability in affect RMSSD across time compared with mean and variance (see the "Descriptive Statistics of State-Specific Affect Through Time" section) suggested promise in leveraging signatures that capture the dynamics of affect expression to ultimately integrate the circumplex model of affect with sentiment analysis. Where observed differences in RMSSD across states argued for potential utility in modeling day-to-day volatility of affect, observed differences in RMSSD across time within any 1 state argued for potential utility in modeling the course of this affect volatility. The circumplex, while not interrogating volatility per se, nevertheless provides a practical, theory-guided means with which *patterns* in written sentiment can be quantified and further investigated within a temporal modeling framework. As such, several additional features were derived from the daily normalized average values of valence (*V*), arousal (*A*), and dominance (*D*) for each state's news headlines calculated in the "Sentiment Analysis" section. These features served as key variables in subsequent modeling efforts (see the "Generalized Mixed Effects Modeling" section) to uniquely operationalize affective variability in VA-2D and VAD-3D circumplex space and ultimately interrogate the significance of daily patterns in sentiment expression as associated with online mental health search behavior outcomes (see the "Mental Health Search Activity" section).

Flux

An operationalization of affect consistency along a dimensional axis of the affect circumplex, flux is the SD of average Z-standardized word affect scores ($Flux_V$, $Flux_D$, $Flux_A$) across all mapped words pooled from all state-specific news story headlines for the day. Flux is scalar—it focuses on intensity and is agnostic to directionality of the word vector in circumplex coordinate space.

(3)

where  is the mean standardized affect scores across all words, w_1, w_2, w_3, ..., w_n.

Pulse

An operationalization of the consistency in extremity (distance from the origin), pulse is the SD of the word vector magnitudes in 2D circumplex space ($Pulse_{VA}$, $Pulse_{VD}$, $Pulse_{AD}$). Pulse is similar in mathematical form (an SD) to flux, except that the SD is found on a set of vector magnitudes rather than scalar quantities of affect—unlike flux, pulse considers both intensity and direction. The vector magnitude, *M*, for a word, *w*, was operationalized in the circumplex as the Euclidean distance in Cartesian coordinate space from the origin (0,0) to (V_w, A_w) in the VA circumplex, to (V_w, D_w) in the VD circumplex, and to (A_w,D_w) in the AD circumplex.

(**4**)

where  is the mean of the vector magnitudes across all words modeled in the circumplex.

Spin

An operationalization of the consistency in angular orientation or the breadth of coverage across the affect space ($Spin_{VA}$, $Spin_{VD}$, $Spin_{AD}$), spin is the SD of the word vectors characterized in terms of their angular displacement from the horizontal 2D circumplex axis (0°). Spin is the opposite of flux in that it is only concerned with position and is therefore intensity agnostic. Like both flux and pulse, spin is at its core a calculation of an SD; however, it is the SD of vector angles (from the x-axis) rather than vector magnitudes or scalar affect values. Thus, the angle formed between the word affect vector and the x-axis of the circumplex, θ, for a word, *w*, is equivalent to the cosine of the dot product of vectors $u = (a_x, 0)$ and $v = (a_x, a_y)$, where a_x and a_y are the x-dimensional and y-dimensional affect values (*V*, *A*, or *D*) associated with the 2D circumplex of interest, respectively, divided by the product of their magnitudes:

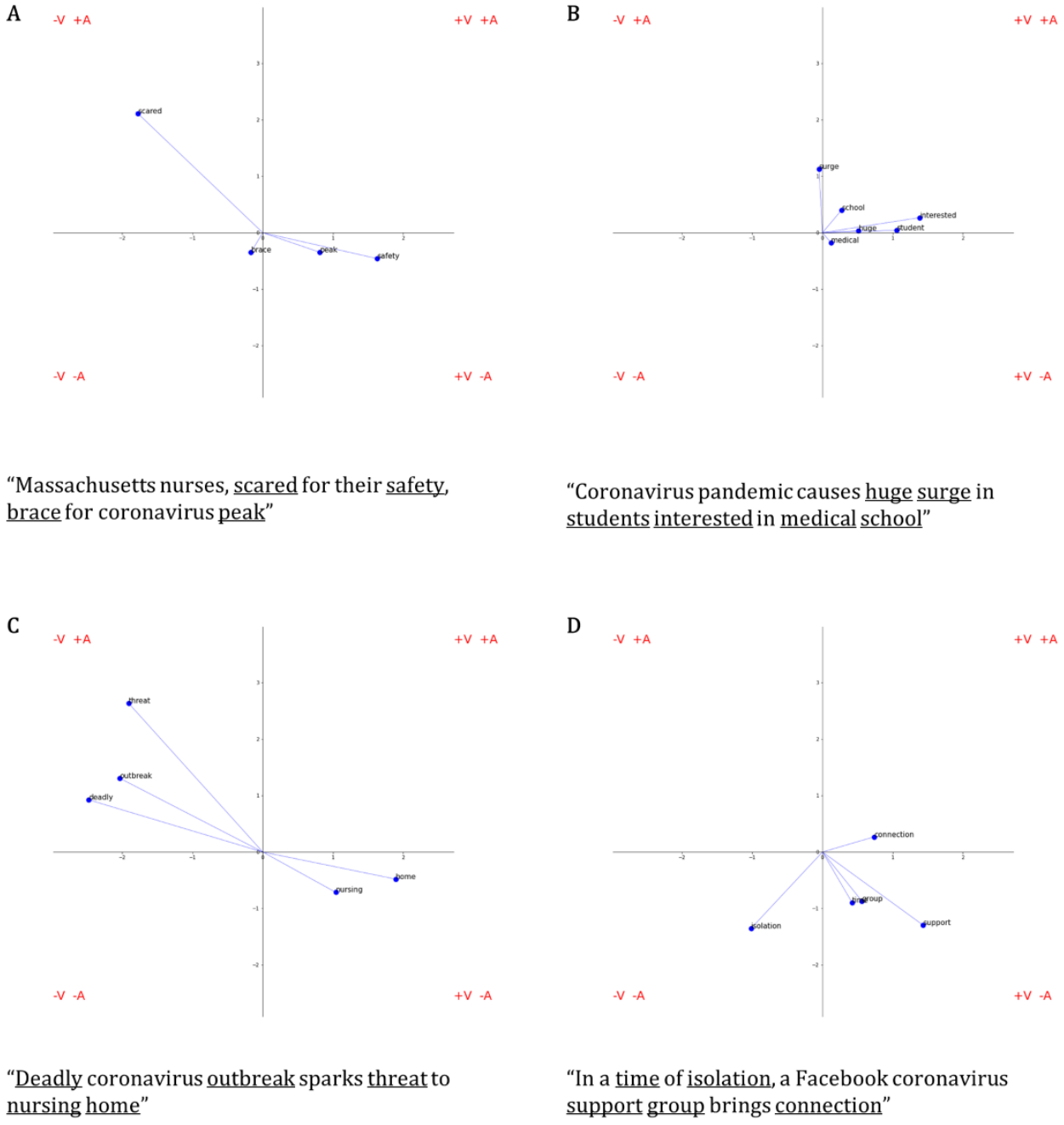
$$\cos\theta_w = (u \cdot v)/||(u \cdot v)|| \quad (5)$$

The final SD equation for calculating spin, with  as the mean of the angular displacement calculations across all words modeled in the circumplex, is therefore:

(**6**)

To provide summative visualizations of descriptive patterns in news headline word choice within the circumplex framework, VA-2D word clouds were derived for each of the 6 representative states from the "Descriptive Statistics of State-Specific Affect Through Time" section (Arizona, Georgia, Illinois, New Jersey, Texas, and Washington). While the efforts outlined in the aforesaid section focused on the deconstruction of basic affect parameters at distinct phases of the evolving pandemic, the efforts herein served 2 purposes: (1) visually introduce the more nuanced circumplex and (2) more broadly characterize word usage among states in a time-agnostic manner. To develop some intuition, Figure 1 provides simple descriptive examples of affect flux and spin profiles from singular Massachusetts news headlines. In practice, however, the resulting circumplex plots represented the most frequently mapped words (≥2 SDs above the mean count of words across news headlines in that state) from a state-wise concatenation of all news headlines across time.

Figure 1. Example valence-arousal affect circumplex plots for COVID news headlines. Headline words that map to the affect dictionary are plotted in affect space with the x-axis representative of a word's valence and the y-axis representative of a word's arousal. The units of both axes represent the Z-score of the word; thus, they can be interpreted as standard deviations from the mean affect score across all words in the reference affect dictionary. Within this derived coordinate space, the circumplex properties of pulse and spin can be intuitively visualized through the mapping of example news headlines. (A) Higher spin and higher pulse. (B) Lower spin and higher pulse. (C) Higher spin and lower pulse. (D) Lower spin and lower pulse.



Generalized Mixed Effects Modeling

To investigate the association between mental health search term activity and affective dynamics of COVID-19 news reporting across the United States, 8 separate mixed effects models were constructed. Because both the VA-2D and VAD-3D operationalizations of affect are popular, the 8 models represent 4 pairs of outcome-specific investigations, where each pair consists of 1 model with VA circumplex variables and another model with VAD circumplex variables (see the "Affect Circumplex Feature Engineering and Exploratory Visualization" section). For each model pair, the outcome was operationalized differently to both holistically and more specifically interrogate mental health search behavior. One subset of the selected terms most closely corresponded with the mental health construct of depression, another distinct subset corresponded with anxiety, and a third aligned with both depression and anxiety. Accordingly, the first model, "All Cluster," pooled the daily state-level counts for all 17 mental health search terms into a single sum; the second model, "Depression Cluster," pooled the daily state-level counts for "apathy," "depression," "hopeless," "suicidal," "suicide," "tired," and "worthless" into a single sum; the third model, "Anxiety Cluster," pooled the daily state-level counts for "afraid," "anxiety," "avoiding,"

"restless," "tense," and "worried" into a single sum; and the fourth model, "Nonspecific Cluster," pooled the daily state-level counts for "angry," "insomnia," "irritable," and "scattered." Given that these outcomes were overdispersed count data (ie, with the variance of the outcome distributions higher than the mean), models were constructed as negative binomial mixed models in R using the *glmmTMB* package (version 1.0.2.1) [48]:

$$\text{MHClusterCount}_{VA} = t + V + A + \text{Spin}_{VA} + \text{Pulse}_{VA} + \text{Flux}_{V} + \text{Flux}_{A} + (1 \mid \text{State}) \quad (7)$$

$$\text{MHClusterCount}_{VAD} = t + V + A + D + \text{Spin}_{VA} + \text{Spin}_{VD} + \text{Spin}_{AD} + \text{Pulse}_{VA} + \text{Pulse}_{VD} + \text{Pulse}_{AD} + \text{Flux}_{V} + \text{Flux}_{A} + \text{Flux}_{D} + (1 \mid \text{State}) \quad (8)$$

For both the VA-2D (Equation 7) and VAD-3D (Equation 8) models, time (day count since March 24, 2020), *t*, and each of the respective affect values and circumplex features were included as fixed effects, while state was included as a random effect. All negative binomial models were run on stacked format tabular data by specifying "family=nbinom2" with default settings in *glmmTMB*. Model fit was assessed using Efron pseudo-r^2 [49]. See Multimedia Appendix 5 for the complete data set used in analysis. For closer intuitive introspection of marginal and interaction effects, prediction plots of the most statistically significant (P<.001) variables were constructed using the *sjPlot* package (version 2.8.7) in R [50].

Ethics

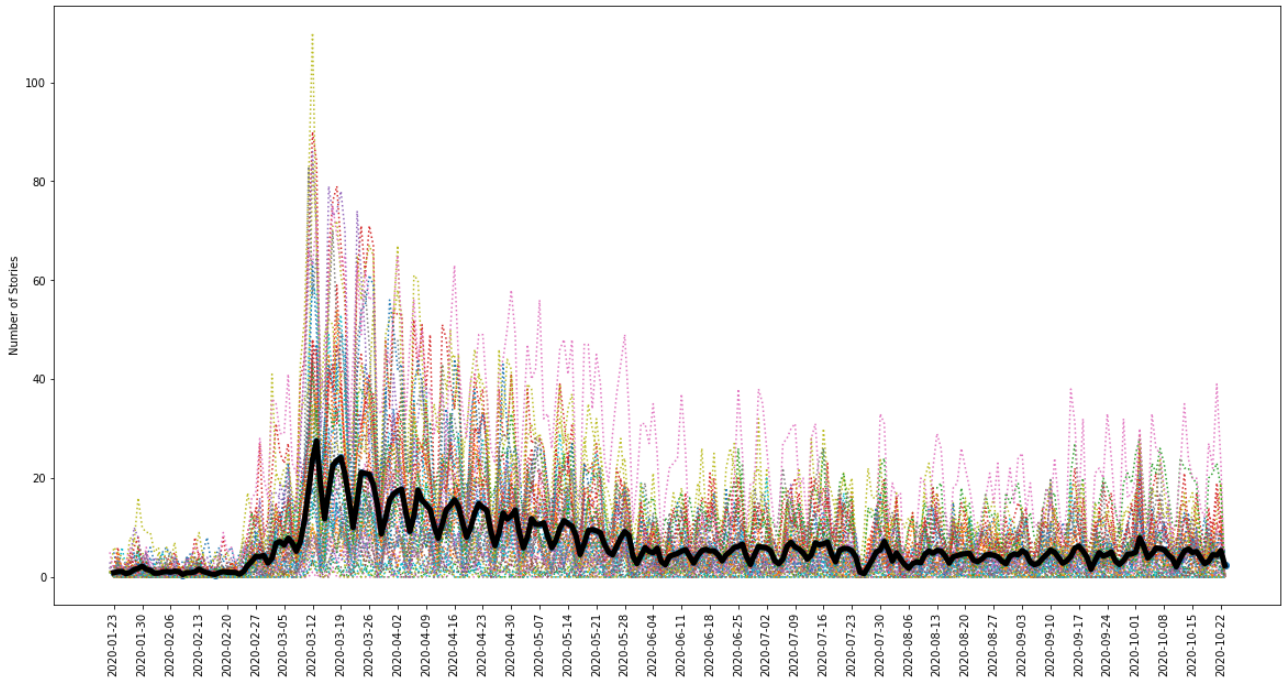
This paper was not considered human subjects research because it used anonymous, publicly available data and as such was exempt from human subjects approval.

Results

Study Sample

Figure 2 illustrates the patterns in overall and state-specific COVID-19 news availability in the current data set. General trends reflect the progression of the pandemic in the United States, with the highest representative coverage occurring in mid-to-late March 2020, coinciding with the official US presidential announcement of a national emergency and the first wave of state-specific announcements and enactments of stay-at-home orders that shortly followed [21]. For a complete list of the outlets utilized by state, along with their respective details on publication city, web traffic, and state circulation rankings, as well as the number of stories included in the data set, please refer to Multimedia Appendix 1.

Figure 2. Trend in available number of COVID-19 news stories across the United States. Colored dashed lines represent state-specific trends in available COVID-19-related news stories through time as determined by the accessibility of the selected news outlets' story headlines via Media Cloud. The bold black line indicates the trend in the average number of representative stories across states. The maximum average number of stories (27.5) occurs on March 13, 2020, coinciding with the release of the presidential Proclamation 9994 that declared COVID-19 a national emergency.

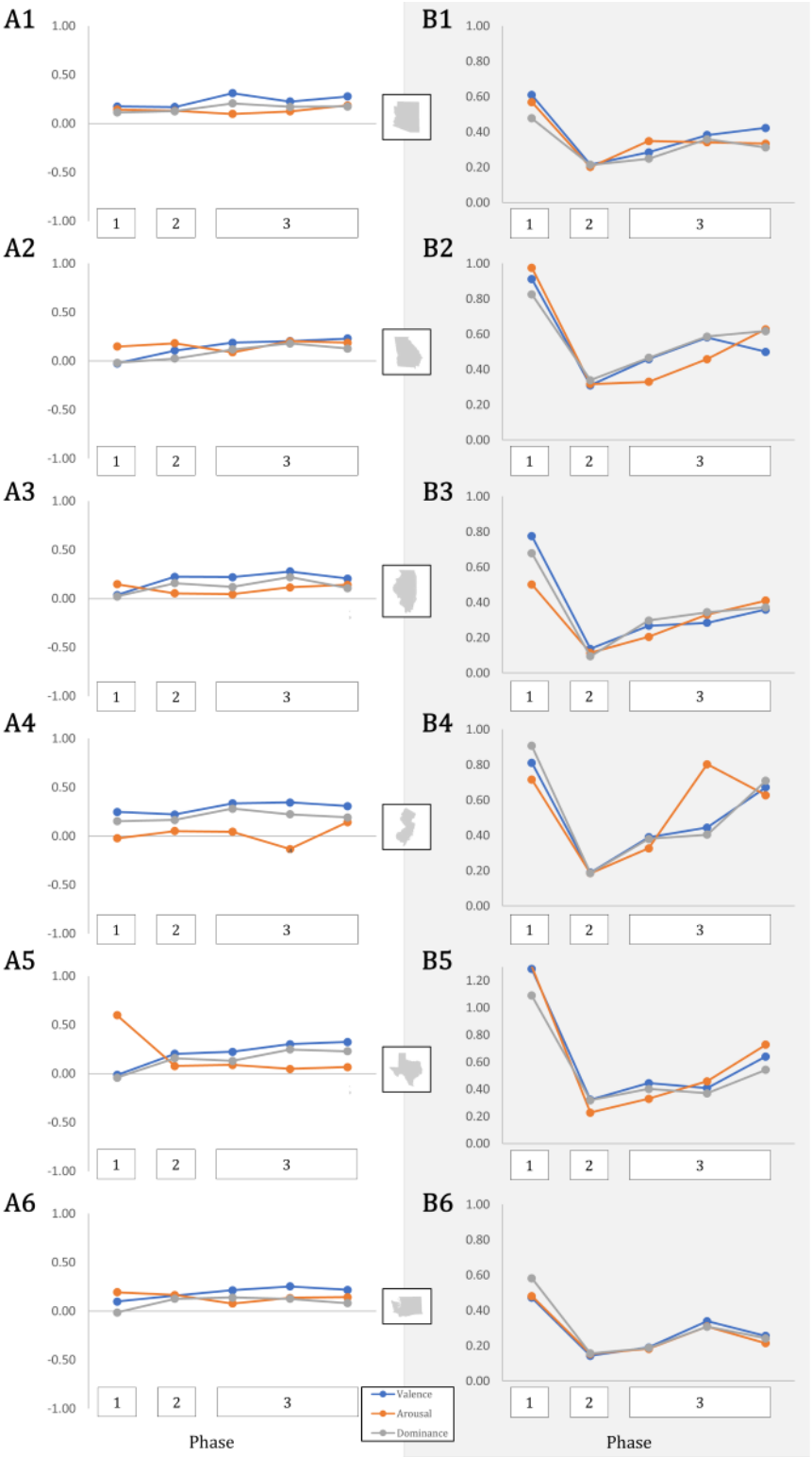


Descriptive Statistics of State-Specific Affect Through Time

Phase-based calculations of standardized affect mean, variance, and RMSSD for each of the 50 states' news headlines revealed several interesting features of the data. In general, the summative phase metrics of mean and variance for each of valence, arousal, and dominance did not exhibit any meaningful trends across phases. For example, all 3 measures of affect tended to hover between a mean standardized score of 0.0 and 0.5, indicating an overall propensity toward slightly positive, arousing, and dominant language in COVID-19–related news headlines in each phase. However, the RMSSD, which captured the change in day-to-day affect over time within each phase, exhibited a general pattern across phases that was consistent in a majority of states. This pattern is clearly illustrated in the 6 representative states of Figure 3B(1-6) and can be described as low consistency (higher RMSSD) in affect during Phase 1, followed by an

appreciable increase in consistency (lower RMSSD) from Phase 1 to Phase 2, leading ultimately to a gradual "rubber-banding" back toward lower consistency across the 198 days of Phase 3. This pattern was observed regardless of the affect dimension in question. Discrepancies between states manifested in terms of the overall magnitude in these shifts, including the extent of "deterioration" in headline affect consistency during Phase 3, thus indicating that despite general similarities in news affect across the nation, local news sentiment over time at different phases of the pandemic is idiosyncratic to the state in question. In turn, this suggests that a fuller appreciation of affect differences in news headlines requires operationalizations that emphasize nuanced *patterns* of expression.

Figure 3. Affect trajectories of representative states across phases of the COVID-19 pandemic. The (A) mean and (B) RMSSD of phase-specific standardized valence (blue), arousal (orange), and dominance (green) across all words comprising COVID-19-related news headlines for each of (1) Arizona, (2) Georgia, (3) Illinois, (4) New Jersey, (5) Texas, and (6) Washington. RMSSD: root mean square of successive differences.



To further illustrate these trajectories of RMSSD in news affect across the country, Figure 4 presents a 3 × 5 grid of percentile-transformed RMSSD choropleths for valence, arousal, and dominance across the 5 predefined time windows of the study. A comparison of affect mean, affect variance, and affect RMSSD in Figures 3 and 5 further emphasizes descriptive statistical similarities and differences across time and state for the 6 representative states. For additional reference, Multimedia Appendix 4 provides all raw statistical results from which time-based descriptions of affect were derived.

Figure 4. Choropleths of news affect percentile ranked RMSSD across phases of the COVID-19 pandemic. The percentile rank of state-specific arousal (green), valence (blue), and dominance (purple) RMSSD across three phases (five temporal partitions) of the pandemic. A pattern of high RMSSD during Phase 1, low RMSSD during Phase 2, and a gradual increase in RMSSD throughout Phase 3 is characteristic of the majority of states. As shown, the temporal window of the study (January 23, 2020 - October 22, 2020) represents the first nine months of the pandemic. RMSSD: root mean square of successive differences.

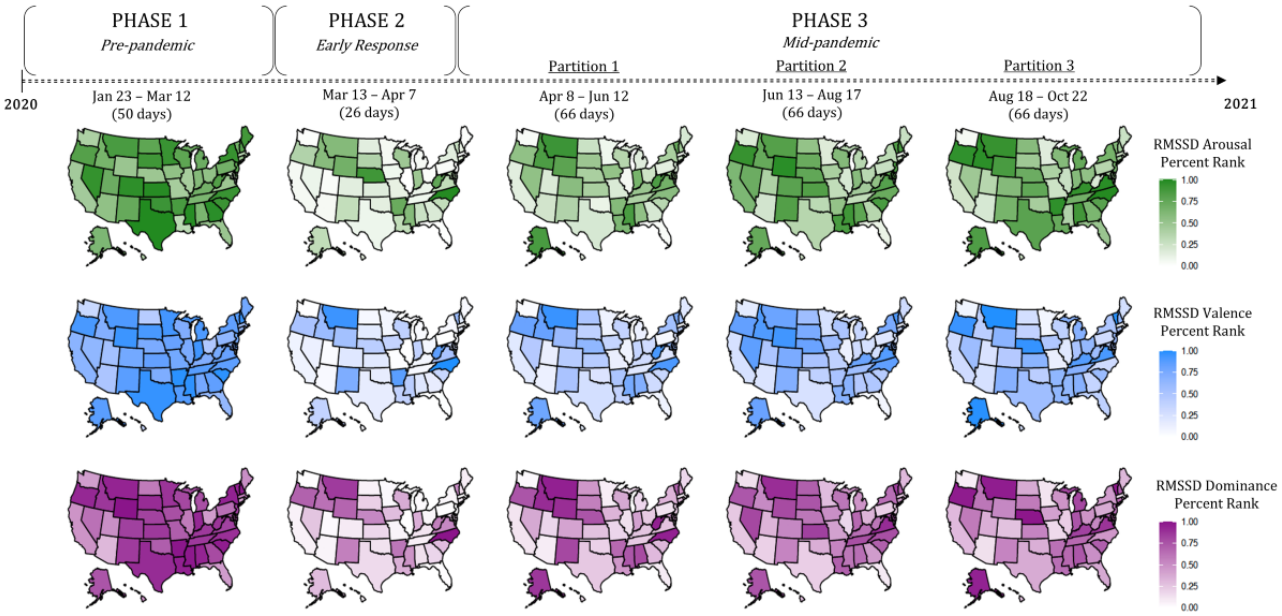
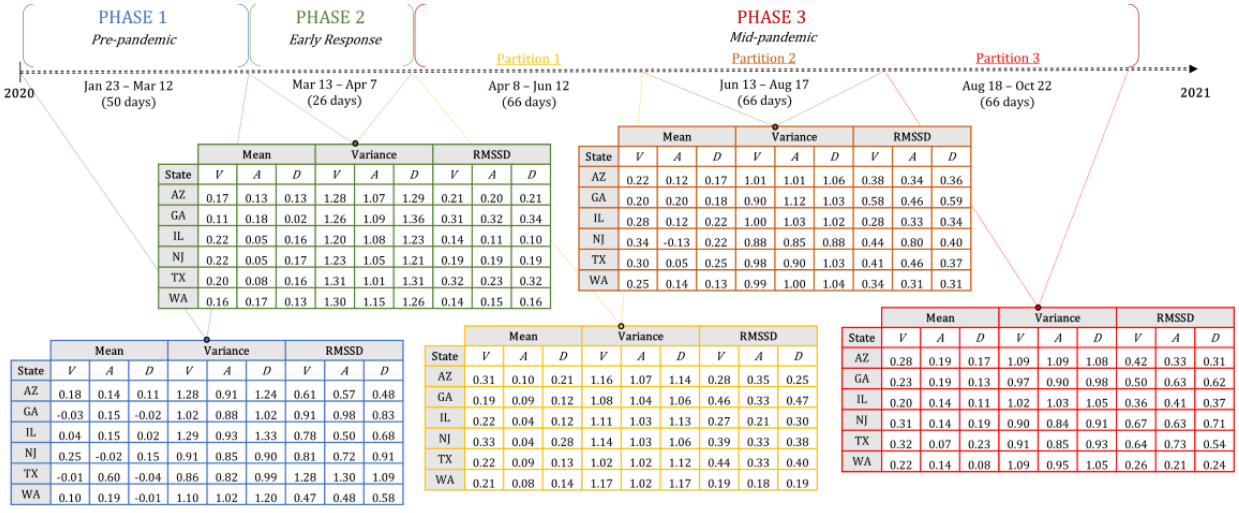


Figure 5. Affect descriptive statistics of representative states across phases of the COVID-19 pandemic. Mean, variance, and RMSSD of daily averaged standardized valence, arousal, and dominance were calculated across news headlines in 6 geographically disparate states. News data were split into 3 distinct phases (with 6 total windows for analysis). A: arousal; D: dominance; RMSSD: root mean square of successive differences; V: valence.



State	Mean			Variance			RMSSD		
	V	A	D	V	A	D	V	A	D
AZ	0.17	0.13	0.13	1.28	1.07	1.29	0.21	0.20	0.21
GA	0.11	0.18	0.02	1.26	1.09	1.36	0.31	0.32	0.34
IL	0.22	0.05	0.16	1.20	1.08	1.23	0.14	0.11	0.10
NJ	0.22	0.05	0.17	1.23	1.05	1.21	0.19	0.19	0.19
TX	0.20	0.08	0.16	1.31	1.01	1.31	0.32	0.23	0.32
WA	0.16	0.17	0.13	1.30	1.15	1.26	0.14	0.15	0.16

State	Mean			Variance			RMSSD		
	V	A	D	V	A	D	V	A	D
AZ	0.22	0.12	0.17	1.01	1.01	1.06	0.38	0.34	0.36
GA	0.20	0.20	0.18	0.90	1.12	1.03	0.58	0.46	0.59
IL	0.28	0.12	0.22	1.00	1.03	1.02	0.28	0.33	0.34
NJ	0.34	-0.13	0.22	0.88	0.85	0.88	0.44	0.80	0.40
TX	0.30	0.05	0.25	0.98	0.90	1.03	0.41	0.46	0.37
WA	0.25	0.14	0.13	0.99	1.00	1.04	0.34	0.31	0.31

State	Mean			Variance			RMSSD		
	V	A	D	V	A	D	V	A	D
AZ	0.18	0.14	0.11	1.28	0.91	1.24	0.61	0.57	0.48
GA	-0.03	0.15	-0.02	1.02	0.88	1.02	0.91	0.98	0.83
IL	0.04	0.15	0.02	1.29	0.93	1.33	0.78	0.50	0.68
NJ	0.25	-0.02	0.15	0.91	0.85	0.90	0.81	0.72	0.91
TX	-0.01	0.60	-0.04	0.86	0.82	0.99	1.28	1.30	1.09
WA	0.10	0.19	-0.01	1.10	1.02	1.20	0.47	0.48	0.58

State	Mean			Variance			RMSSD		
	V	A	D	V	A	D	V	A	D
AZ	0.31	0.10	0.21	1.16	1.07	1.14	0.28	0.35	0.25
GA	0.19	0.09	0.12	1.08	1.04	1.06	0.46	0.33	0.47
IL	0.22	0.04	0.12	1.11	1.03	1.13	0.27	0.21	0.30
NJ	0.33	0.04	0.28	1.14	1.03	1.06	0.39	0.33	0.38
TX	0.22	0.09	0.13	1.02	1.02	1.12	0.44	0.33	0.40
WA	0.21	0.08	0.14	1.17	1.02	1.17	0.19	0.18	0.19

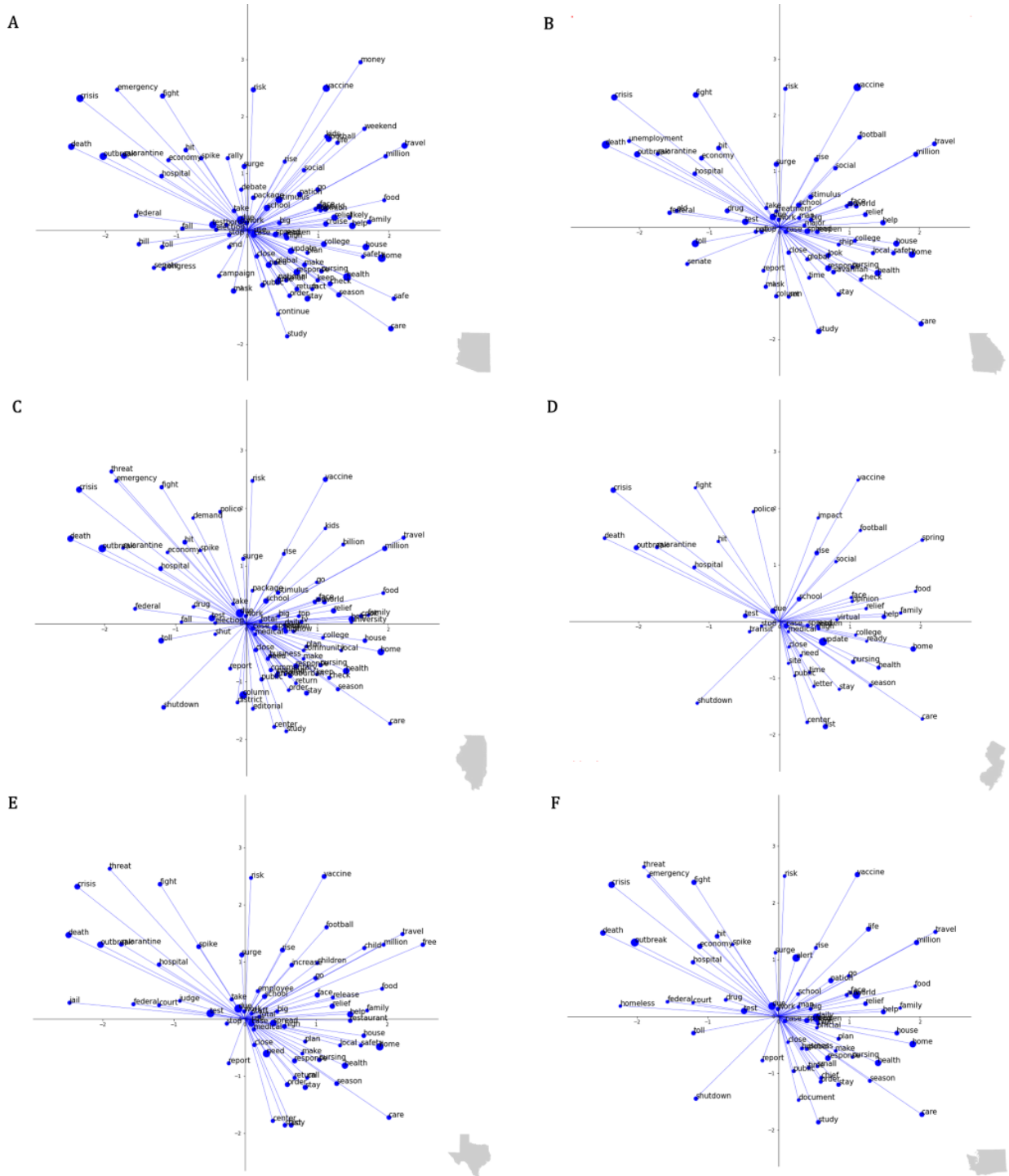
State	Mean			Variance			RMSSD		
	V	A	D	V	A	D	V	A	D
AZ	0.28	0.19	0.17	1.09	1.09	1.08	0.42	0.33	0.31
GA	0.23	0.19	0.13	0.97	0.90	0.98	0.50	0.63	0.62
IL	0.20	0.14	0.11	1.02	1.03	1.05	0.36	0.41	0.37
NJ	0.31	0.14	0.19	0.90	0.84	0.91	0.67	0.63	0.71
TX	0.32	0.07	0.23	0.91	0.85	0.93	0.64	0.73	0.54
WA	0.22	0.14	0.08	1.09	0.95	1.05	0.26	0.21	0.24

Temporally Agnostic Exploratory Visualization With the Affect Circumplex

With a focus on the VA-2D operationalization of affect dynamics, the resulting circumplex word clouds in Figure 6 highlight interesting time-agnostic similarities and differences in COVID-19–related news headlines across the country. Most generally, descriptive analysis of the 6 representative states indicated an overrepresentation of neutral and positively valenced diction (eg, “health”) with some state-specific variety. In addition, more negatively valenced and strongly arousing language (eg, “emergency”) was less prevalent, however more consistent and ubiquitous across states. The results, in conjunction with those in the “Descriptive Statistics of State-Specific Affect Through Time” section, encourage the

implementation of affect-based models that account for state-specific random effects within a temporal framework.

Figure 6. Valence–arousal circumplex of most commonly used words in COVID-19 news headlines across representative states. Valence–arousal circumplex plots of the most frequently used words in COVID-19–related news headlines. Plots are anchored in 2D Euclidean space where the x-axis indicates standardized valence of a word and the y-axis indicates standardized arousal of a word. All represented words reflect a state-specific frequency of use at or above 2SDs from the mean use across all words in the news headlines for the respective state. The size of the data point scales with relative frequency. Words used more frequently are represented by larger points relative to those used less frequently. (A) Arizona, (B) Georgia, (C) Illinois, (D) New Jersey, (E) Texas, and (F) Washington.



Generalized Mixed Effects Modeling

Eight different negative binomial mixed effects models (in VA-2D and VAD-3D pairs) were constructed to interrogate the associations between affect dynamics in COVID-19–related news and online mental health–related search term activity. Activity was operationalized in terms of 4 clusters. First, the VA-2D (pseudo-r^2=0.11) and VAD-3D (pseudo-r^2=0.11) models corresponding to the "All Cluster" of mental health search terms corroborated a statistical significance in the fixed effects of both $Spin_{VA}$ (P=.004 and .009, respectively) and $Flux_A$

(P<.001 and P<.001, respectively) with similar respective magnitudes of association (Table 1). Specifically, $Spin_{VA}$ was found to have a significantly (P=.004) negative association with mental health search term volume: a unit increase in $Spin_{VA}$ was found to have an approximate 14.8% decrease in mental health search term volume. By contrast, $Flux_A$ was found to have a significantly (P<.001) positive association with mental health search term volume: a unit increase in $Flux_A$ reflects an approximate 24.7% increase in mental health search term counts. For completeness, Multimedia Appendix 6 reports on the conditional modes of the state-based random effects for these models.

Table 1. Generalized mixed effects model results for all cluster of mental health search terms[a].

Fixed effects feature	VA[b] model			VAD[c] model		
	β	95% CI	*P* value	β	95% CI	*P* value
Time (*t*)	–.013	–0.035 to 0.009	.26	–.014	–0.036 to 0.009	.23
Valence (*V*)	–.061	–0.126 to 0.004	.07	–.047	–0.136 to 0.043	.31
Arousal (*A*)	–.01	–0.077 to 0.057	.77	–.015	–0.082 to 0.053	.67
Dominance (*D*)				–.034	–0.119 to 0.052	.44
$Spin_{VA}$	–.16	–0.270 to –0.050	.004[d]	–.17	–0.297 to –0.042	.009[d]
$Spin_{VD}$				.01	–0.103 to 0.123	.87
$Spin_{AD}$				.03	–0.092 to 0.151	.63
$Pulse_{VA}$	.131	–0.031 to 0.292	.11	.095	–0.108 to 0.299	.36
$Pulse_{VD}$				–.314	–0.717 to 0.089	.13
$Pulse_{AD}$				.035	–0.179 to 0.249	.75
$Flux_V$	.016	–0.084 to 0.115	.76	.318	–0.006 to 0.642	.05
$Flux_A$	.221	0.116 to 0.326	<.001[e]	.216	0.106 to 0.325	<.001[e]
$Flux_D$				.09	–0.231 to 0.411	.58

[a]Results of the negative binomial mixed effects model for the "All Cluster" summed mental health terms outcome. The VA model includes affect and circumplex features representing the valence and arousal dimensions of words, while the VAD model includes affect and circumplex features representing the valence, arousal, and dominance dimensions of words.

[b]VA: valence–arousal.

[c]VAD: valence–arousal–dominance.

[d]P<.05.

[e]P<.001.

Second, the VA-2D (pseudo-r^2=0.11) and VAD-3D (pseudo-r^2=0.11) models corresponding to the "Anxiety Cluster" of mental health search terms suggest a consistent statistically significant (P<.001 and P<.001, respectively) positive association between $Flux_A$ and anxiety-related search term volume, with $Flux_V$ (P=.04) and D (P=.03) exhibiting additional significance in the VAD-3D model only (Table 2). Significance (P<.001) in $Flux_A$ partially echoes findings in the "All Cluster" with a 1 unit increase in $Flux_A$ accounting for an approximate 24.5% increase in anxiety-related online search activity. Idiosyncratic to the VAD-3D model, a 1 unit increase in $Flux_V$ was associated with an approximately 42.0% increase, while a 1 unit increase in D was associated with an approximately 9.1% decrease in anxiety-related search term counts. For completeness, Multimedia Appendix 6 reports on the conditional modes of the state-based random effects for these models.

Table 2. Generalized mixed effects model results for anxiety cluster of mental health search terms[a].

Fixed effects feature	VA[b] model			VAD[c] model		
	β	95% CI	*P* value	β	95% CI	*P* value
Time (*t*)	–0.02	–0.042 to 0.003	.08	–0.023	–0.045 to 0.000	.05
Valence (*V*)	–0.06	–0.126 to 0.005	.07	–0.003	–0.094 to 0.087	.94
Arousal (*A*)	–0.016	–0.084 to 0.052	.65	–0.025	–0.094 to 0.044	.48
Dominance (*D*)				–0.095	–0.181 to –0.008	.03[d]
$Spin_{VA}$	–0.109	–0.220 to 0.002	.06	–0.11	–0.239 to 0.189	.09
$Spin_{VD}$				0.007	–0.109 to 0.123	.90
$Spin_{AD}$				0.036	–0.088 to 0.160	.57
$Pulse_{VA}$	0.078	–0.085 to 0.241	.35	0.032	–0.173 to 0.238	.76
$Pulse_{VD}$				–0.357	–0.768 to 0.053	.09
$Pulse_{AD}$				0.038	–0.178 to 0.254	.73
$Flux_{V}$	0.006	–0.094 to 0.106	.91	0.351	0.021 to 0.680	.04[d]
$Flux_{A}$	0.219	0.112 to 0.326	<.001[e]	0.215	0.104 to 0.327	<.001[e]
$Flux_{D}$				0.092	–0.235 to 0.419	.58

[a]Results of the negative binomial mixed effects model for the "Anxiety Cluster" summed mental health terms outcome. The VA model includes affect and circumplex features representing the valence and arousal dimensions of words, while the VAD model includes affect and circumplex features representing the valence, arousal, and dominance dimensions of words.

[b]VA: valence–arousal.

[c]VAD: valence–arousal–dominance.

[d]*P*<.05.

[e]*P*<.001.

Third, the VA-2D (pseudo-r^2=0.10) and VAD-3D (pseudo-r^2=0.10) models corresponding to the "Depression Cluster" of mental health search terms uncovered consistent statistical significance for the fixed effects of $Spin_{VA}$ (*P*<.001 and *P*<.001, respectively) and $Flux_{A}$ (*P*<.001 and *P*<.001, respectively) with additional statistical significance for *V* (*P*=.04) and $Pulse_{VA}$ (*P*=.03) within the VA-2D model only (Table 3). Akin to what was observed in the "All Cluster", $Spin_{VA}$ was found to have a significantly (*P*=.003) negative association with depression-related search term counts: a unit increase in $Spin_{VA}$ was associated with an 18.7% decrease in depression-related search volume. Likewise, $Flux_{A}$ had a significantly (*P*<.001) positive association with depression-related search activity: there was an approximate 28.9% increase in depression-related search counts for every 1 unit increase in $Flux_{A}$. Idiosyncratic to the VA-2D model, a 1 unit increase in *V* was associated with an approximate 6.9% decrease, while a 1 unit increase in $Pulse_{VA}$ was associated with an approximate 20.2% increase in depression-related search term counts. For completeness, Multimedia Appendix 6 reports on the conditional modes of the state-based random effects for these models.

Table 3. Generalized mixed effects model results for depression cluster of mental health search terms[a].

Fixed effects feature	VA[b] model			VAD[c] model		
	β	95% CI	*P* value	β	95% CI	*P* value
Time (*t*)	–0.015	–0.037 to 0.007	.18	–0.016	–0.039 to 0.007	.17
Valence (*V*)	–0.071	–0.138 to –0.071	.04[d]	–0.072	–0.164 to 0.020	.13
Arousal (*A*)	–0.009	–0.077 to 0.059	.80	–0.01	–0.079 to 0.059	.78
Dominance (*D*)				–0.012	–0.099 to 0.076	.80
$Spin_{VA}$	–0.207	–0.320 to –0.094	<.001[e]	–0.228	–0.358 to –0.098	<.001[e]
$Spin_{VD}$				0.015	–0.101 to 0.130	.80
$Spin_{AD}$				0.048	–0.078 to 0.173	.45
$Pulse_{VA}$	0.184	0.017 to 0.351	.03[d]	0.138	–0.072 to 0.348	.20
$Pulse_{VD}$				–0.247	–0.657 to 0.164	.24
$Pulse_{AD}$				0.052	–0.170 to 0.273	.65
$Flux_{V}$	0.049	–0.053 to 0.152	.35	0.301	–0.031 to 0.633	.08
$Flux_{A}$	0.254	0.146 to 0.362	<.001[e]	0.246	0.133 to 0.358	<.001[e]
$Flux_{D}$				0.041	–0.287 to 0.368	.81

[a]Results of the negative binomial mixed effects model for the "Depression Cluster" summed mental health terms outcome. The VA model includes affect and circumplex features representing the valence and arousal dimensions of words, while the VAD model includes affect and circumplex features representing the valence, arousal, and dominance dimensions of words.

[b]VA: valence–arousal.

[c]VAD: valence–arousal–dominance.

[d]*P*<.05.

[e]*P*<.001.

At the request of an anonymous reviewer, identical negative binomial mixed effects models with time modeled as logistic were run to challenge the above models' assumption of time as linearly associated with the outcome. Using Efron pseudo-r^2 to assess comparative fit of the logistic-transformed time models with their respective linear version, results were comparable to 2 decimal places (pseudo-r^2=0.10-0.11). As such, treating time as linear or logistic did not lead to an appreciable difference in fit.

Both spin and flux were consistently found to be significant (*P*<.05) in the above mixed models. Given this prominence, Figure 7 first provides an intuitive visualization of the marginal effect of the cluster-/model-specific significant (*P*<.05) spin and flux-based parameters on the frequency of mental health search term clusters, indicating that lower spin and higher flux individually result in larger predicted counts of cluster-related search terms. Considering the significance of $Spin_{VA}$ and $Flux_{A}$ simultaneously, Figure 8 then illustrates their interactive effect on the predicted summative count of all mental health search terms (All Cluster), further implicating overall mental health search term behavior as positively associated with $Flux_{A}/Flux_{V}$ and negatively associated with $Spin_{VA}$.

Figure 7. Marginal effect plots of significant VA-2D affect circumplex features on mental health search term counts. The prediction plots illustrate the marginal impact of the most significant affect features on online mental health search term count outcome from the cluster's respective VA-2D negative binomial mixed model. The x-axis indicates the potential value (0-3) of the designated affect feature, and the y-axis is the model-predicted search term count. Slopes recapitulate the directionalities of association as enumerated in Tables 1-3. (A) "All Cluster"; (B) "Anxiety Cluster"; (C) "Depression Cluster". VA: valence–arousal. VA-2D: valence-arousal-two-dimensional.

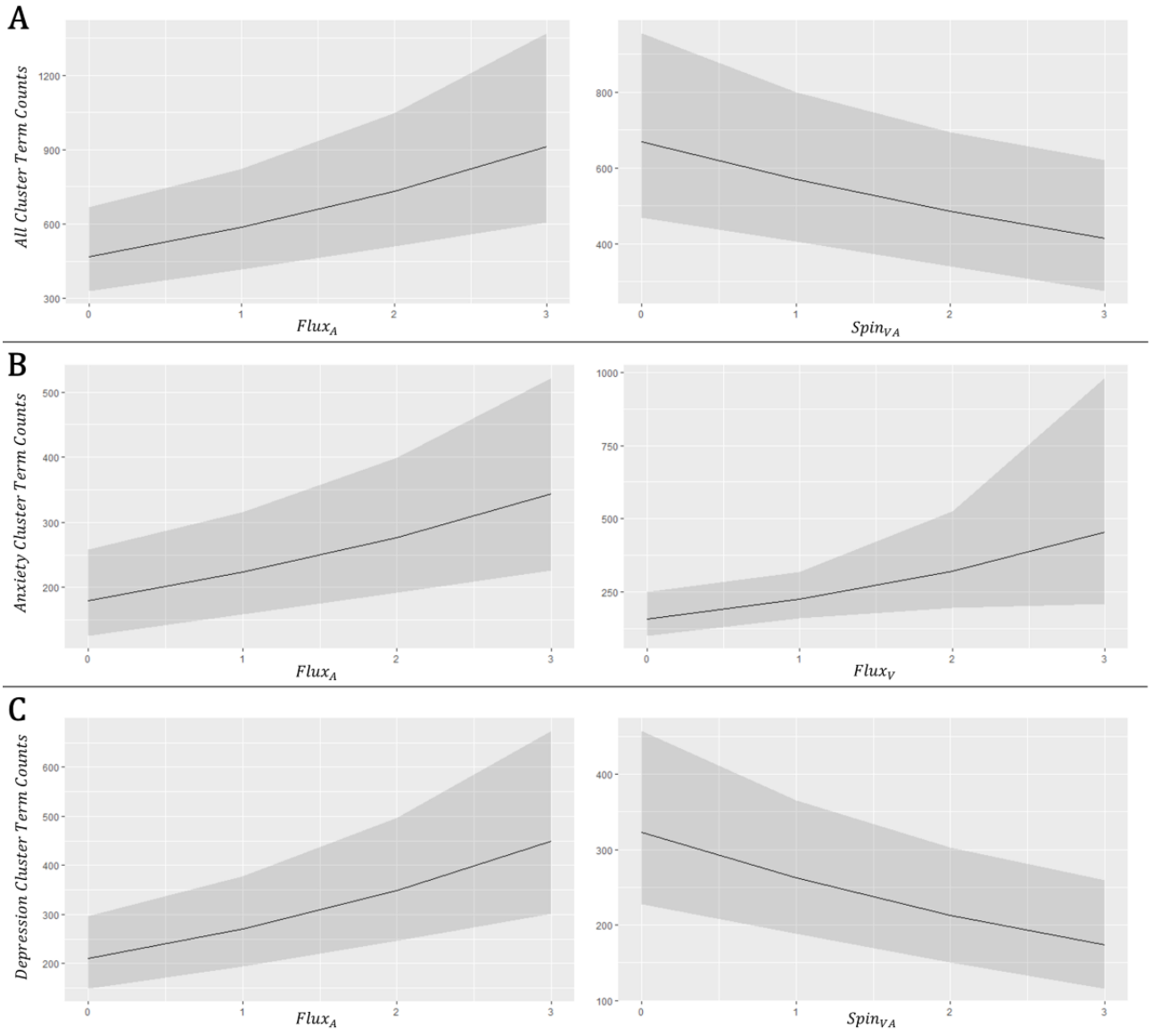
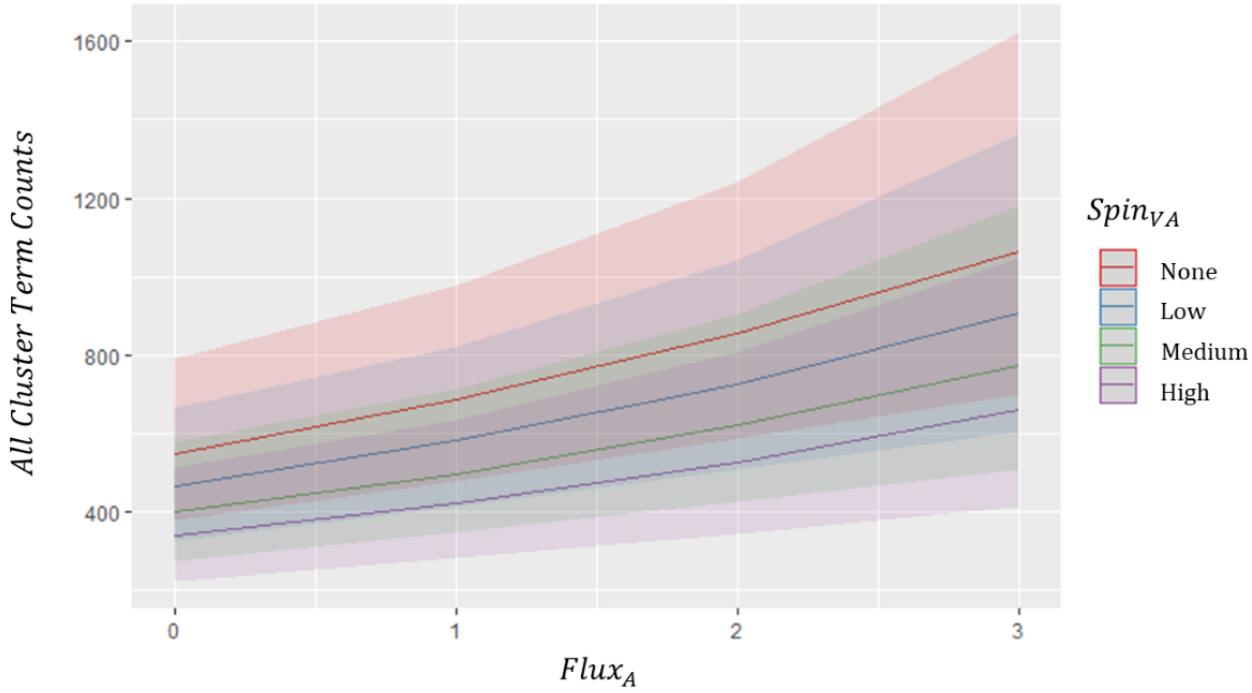


Figure 8. Interaction plot of overall significant affect circumplex features from VA-2D All Cluster model. Illustration of the marginal interactions of the most significant circumplex features from the fit VA-2D circumplex "All Cluster" negative binomial mixed model. Higher $Flux_A$ influences predictions toward higher overall activity in online mental health search term counts, while simultaneous higher $Spin_{VA}$ dampens, but does not eliminate, the overall trajectory toward higher search term activity. VA: valence–arousal. VA-2D: valence-arousal-two-dimensional.



Discussion

This research endeavor revealed patterns in day-to-day COVID-19 news headline affect variation across the first 9 months of the pandemic. Applying a rule-based, empirical approach to sentiment analysis, time-agnostic similarities and differences in word use and affect were found across the United States, including the ubiquitous use of negatively valenced and strongly arousing language. Initial temporal analyses revealed trends in affect change within and between phases of the pandemic, highlighting the potential importance of operationalizing patterns of expression in future investigations involving written sentiment. For example, the generalized mixed-effects models employed in this work utilized circumplex-based predictors, revealing that an increased consistency in affective tone ($Spin_{VA}$) was predictive of increased depression-related search term activity, and emotional language patterns indicative of affective uncontrollability ($Flux_A$) contributed more generally to an increase in online mental health search term frequency.

This study leveraged both news and internet search data to interrogate the association between the daily affective dimensions of local news headlines and online mental health search behavior across the United States. Using the recent COVID-19 pandemic as an example, this research applied the circumplex theory of affect within the context of NLP to operationalize the dynamics of language-based emotion. These operationalizations were then employed within a time-anchored, generalized mixed modeling framework to gauge the significance of the circumplex properties of news headlines on same-day mental health–related search term volume. The study aimed to first explore news coverage temporally, holistically, and geographically from the unique quantitative perspective of word affect (ie, valence, arousal, and dominance). Through basic descriptive statistics and summative data visualization, these efforts profiled key similarities and differences which informed and justified modeling choices for subsequent hypothesis testing. After establishing this descriptive baseline of COVID-19–related news affect, the research was interested in testing hypotheses concerning the relative significance of affect flux, pulse, and spin to ultimately contribute a unique and more nuanced explanation of the semantic mechanisms that underlie the impact of written news exposure on the mental health of society.

As an initial exploratory analysis, this study took daily averaged values of word use affect in COVID-19 news headlines and calculated the state- and pandemic phase–specific mean, variance, and RMSSD of these values. One primary goal of this endeavor was to compare affect coverage through time from both traditional, pattern-agnostic quantifications (ie, mean and variance) with one that began to interrogate notions of affective dynamics (RMSSD within a phase). The results indicated that while the mean and variance of phase-specific COVID-19 news affect did not show significant change through time (Figure 5), there was a clear recapitulation of RMSSD patterns of change across the vast majority of states (Figure 4). Example comparisons of mean and RMSSD in 6 representative states further reflected the nature of these statistical patterns (Figure 3). Taken together, these findings indicated several important aspects of COVID-19 news headline sentiment.

First, and most broadly, these findings indicate that the phenomenology of written sentiment in the context of news

coverage is volatile and dynamic; RMSSD can be thought of as an inspiration and justification for applying circumplex-based modeling in subsequent hypothesis testing. Second, deviance among the 50 states indicates that while there is a potentially significant pattern of affect change through time, it does not manifest consistently and to the same degree across the United States. Thus, modeling should account for "state" as a random effect when interrogating any affect–outcome relationship from the data, insinuating the utility of a mixed modeling strategy. Third, these results provide a baseline, descriptive statistical narrative of COVID-19 news coverage throughout the first 9 months of the pandemic. The employment of emotional written language in news, while showing little variance or change in central tendency between pandemic phases holistically (Figures 3A and 5), was characterized by appreciable patterns of change *within each phase* which differ from one phase to the next (Figures 3B(1-6), 4, and 5). Thus, day-to-day fluctuations in news sentiment, as well as the patterns of these fluctuations in local news, may be potentially informative and impactful aspects of news coverage on the resulting mental health of the American people. In general, the initial findings supported the notion that written news affect in this setting is nuanced and thus requires operationalizations and associated modeling that can more subtly capture changes in the dimensions of sentiment through time.

Before any digital surrogates of mental health were incorporated for hypothesis testing, this study sought to additionally provide some linguistic characterization to the headline content driving the aforementioned statistical patterns of sentiment. The resulting word clouds depicting the most commonly used words in COVID-19 news headlines across the 6 representative states aimed not only to interrogate the nuances in state-specific diction, but also to consider the valence and arousal of these headlines at the word level (Figure 6). Words with high arousal and low valence (eg, "emergency," "quarantine," and "fight") were ubiquitous across the majority of the representative states, likely reflecting common COVID-19–related headlines that may not be idiosyncratic to state-specific news outlets and therefore representative of common themes in news coverage across the United States. However, the word clouds of the representative states predominantly consisted of words with high valence and varying arousal. While some of these words suggest a direct connection to COVID-19–related updates (eg, "vaccine," "travel," and "stimulus"), others are more indicative of the impact of COVID-19 on everyday life specific to that state (eg, "football," "college," and "restaurant"). The over-representation of words with high valence is worthy of note as it may lend to a misinterpretation of COVID-19 news headlines as being generally positive, as reflected by a positive mean valence across the top most commonly used words in a given state. This interpretation highlights the necessity for contextualization of the headline's diction. For example, when mapped against the valence–arousal circumplex, the headline "Massachusetts nurses, scared for their safety, brace for coronavirus peak" produces 2 words of positive valence ("safety" and "peak"), matched by 2 words of negative valence ("scared" and "brace"; Figure 1A). Despite an inconclusive depiction of the headline's sentiment when operationalized by the mean valence of the 4 individually mapped words, the headline as a whole suggests a clear sense of fear. Such examples highlight that while the incorporation of mean valence may allow for comparison across a single affect dimension, additional introspection is required to better understand temporal and multidimensional variation in news headline sentiment.

Following the exploratory and descriptive endeavors of COVID-19 news headline coverage, the study turned to the core issue at hand, namely, the interrogation of the impact of these daily news headlines on the mental health–related online search behavior of the US population. Using same-day Google Trends count data of key mental health terms, negative binomial mixed effects models were constructed using time and affect circumplex-derived variables. Valence, arousal, and dominance (ie, flux, pulse, and spin) were used as fixed effects in both VA-2D and VAD-3D representations of word affect along with "state" as a random effect (see the "Generalized Mixed Effects Modeling" section). The results of these circumplex-based modeling efforts can be partially understood through the lens of established psychological theory and are concomitant with word affect patterns that reflect the zeitgeist of the early pandemic. In addition, the novel application of sentiment analysis within a circumplex framework has highlighted several novel foci for future research consideration.

One key methodological feature of the hypothesis-driven mixed modeling was the summation of daily state-specific mental health search term count activity by 4 clusters: (1) All, (2) Depression, (3) Anxiety, and (4) Nonspecific (see the "Mental Health Search Activity" and "Generalized Mixed Effects Modeling" sections). For terms related to the construct of depression, the models implicate both spin in the valence–arousal circumplex ($Spin_{VA}$) and flux–arousal ($Flux_{A}$) as statistically significant predictors of depression-related search term activity through time in both the VA-2D and VAD-3D representations of affect (Table 3). This result indicates that both the affective position and magnitude of words in news headline exposure were associated with depression-related search activity. Moreover, this result differed from the results of the "Anxiety" cluster models, where only $Flux_{A}$ was a statistically significant predictor of anxiety-related search activity in both the VA-2D and VAD-3D versions of the model. The exclusive significance of $Flux_{A}$ in the anxiety models implies that only the magnitude of the arousal dimension is important for prediction when anxiety-related terms are considered (Table 2). Additionally, the "All Cluster" models, where anxiety, depression, and nonspecific search terms were considered, reflected the union of circumplex variable significance for the "Anxiety", "Depression", and "Nonspecific" (not-shown; no significance) clusters (Table 1). When considering the importance of spin for depression-related search behavior, $Spin_{VA}$ can be thought of as representing affective "tone" or "mood" (ie, the "kind" of affect [37]) due to its solely position-based (and not magnitude-based) operationalization within the circumplex. Thus, the finding that lower spin (more consistent tone) is associated with higher depression-related search term activity may be a reflection of the mundane, of a situation that is stagnant and inescapable, thereby potentially fueling depressive thoughts and feelings. Such a description echoes feelings of languishing—a void between depression and

peak psychological well-being—that have widely characterized the mental milieu of those affected by the COVID-19 pandemic [51].

Turning now to the significance of $Flux_A$ in the "Anxiety Cluster," "Depression Cluster," and "All Cluster" models, the learned helplessness hypothesis may offer some explanation. The learned helplessness hypothesis describes a situation where perceived uncontrollable events lead to a learned disconnection between behavior and outcome, ultimately producing motivational, cognitive, and emotional states of uncontrollability during this learning process [52]. This framing purports that uncontrollable aversive events produce greater emotional disruption than controllable aversive events, which ultimately leads to what could be described as a "defeatist" mentality (of particular note for depression). The significance of $Flux_A$ as a predictor of mental health search term counts highlights the affective uncontrollability of COVID-19 reporting (and COVID-19 itself); ubiquitous feelings of uncertainty surrounding the pandemic have been parroted by the news, thereby reinforcing these sentiments through time. The positive directionality of association between $Flux_A$ and the "Anxiety Cluster," "Depression Cluster," and "All Cluster" model outcomes is indicative of the impact of greater-expressed uncertainty on the mental health of the population.

In joint consideration of low spin and high flux, with importance centered on valence and arousal, the modeling results showed that COVID-19–related news coverage through time clashes low energy and high energy emotional states, a combination (and fluctuation) of both the arousing and the mundane. Given the behavioral and environmental context and uncertainties surrounding the pandemic, this dynamic may be particularly disturbing for consumers and may engender heightened feelings of both depression and anxiety, ultimately exacerbating mental health issues. Figures 7 and 8 offer some preliminary empirical support for this association.

The application and similarity of results between the VA-2D and VAD-3D models of affect support the overall robusticity of the results. Such a general recapitulation also implies that dominance, in the setting of news headlines, does not seem to contribute to modeling the association between written news affect and mental health search term behavior; the VA-2D circumplex statistically accounts for mental health search term behavior on its own. However, there are some discrepancies worth noting. In modeling anxiety-related search terms, flux valence ($Flux_V$) is significant in the VAD-3D model only, meaning the dynamics of valence become important only after controlling for dominance. This supports a common critique of the 2D circumplex, which states that the limited framing may not adequately capture nuances in emotion. For example, nervousness and anger share a similar space in the VA circumplex (negative valence, positive arousal), but with the addition of dominance in the VAD circumplex, nervousness (negative dominance) and anger (positive dominance) become separable in 3D space. While the current results saw limited instances of discrepancy between the 2, it is worth noting for future endeavors that accounting for dominance may uncover nuanced differences in the affective properties and associated circumplex dynamics that exist among words and their corresponding emotional states. Dimensionality notwithstanding, this study has underlined arousal (over both valence and dominance) as the most significant affective dimension in consideration of mental health search behavior outcomes. The implications of the significance of arousal are not clear and should be a focus for future work in this area.

This study contributes to a small yet growing body of literature connecting NLP, sentiment analysis, and affective dynamics, ultimately offering a novel framework to more finely interrogate the impact of emotion in written form. Leveraging a large and robust data set of time-anchored news headlines and internet search behavior, theory-driven operationalizations of emotion, along with empirically informed modeling of association against the backdrop of a widespread emotional and impactful stimulus, this work serves both as a theoretical proof of concept for methodological integration and as a practical, big data–driven investigation into the impact of language on the mental health of society. While acknowledging these strengths, this study also has several limitations worth noting when considering the significance of the results. First and foremost, the analyses only extracted individual words, or unigrams, from news headlines and used a reference dictionary to quantify the sentiment of each word. While considering unigrams is convenient for extracting the sentiment of individual words, they do not take into account contextual elements of full sentences that may disambiguate the meaning and therefore sentiment of a word [53]. For instance, the word "good" can be modified with an intensifier like "very" to increase its positive sentiment or a negator like "not" to give it a negative sentiment [54]. Thus, using an *n*-gram approach, where *n* is a number of words greater than 1 (eg, bigrams, trigrams), may have yielded more accurate quantifications of sentiment contained within the COVID-19–related news headlines. However, because this study utilized an externally developed affect reference dictionary for English lemmas, which mapped unigrams to sentiments, the authors of this study would have had to empirically derive a new reference dictionary for *n*-grams of a comparable quality to the one used. In addition, because the external reference dictionary could only account for 34.47% (252,254/731,607) of the lemmas extracted from the COVID-19–related news headlines, the majority of the extracted lemmas could not be considered for analysis. Second, this study relied on the Media Cloud API client for COVID-19–related headline availability. It is possible that some news stories were not available in the database and therefore not included in this analyses. While the news headline corpus was extensive in totality, state-wise news coverage was dependent on database availability and thus not equally representative nor exhaustive. Finally, the present methods used approximate calculations of mental health–related term frequencies based on normalized Google Trends values, as Google Trends does not provide absolute search term counts. These estimated frequencies were conservative, consistent, and empirically based, and thus the impact of using them on the present results was likely slim; however, the true search term counts may yield slightly different results.

This work demonstrated the general promise of leveraging the circumplex model of affect to written content, providing an

example for how circumplex theory can be integrated with NLP and sentiment analysis techniques. The results of this work operationalized sentiment analysis via the theory-driven affect circumplex to uncover nuanced dynamics of word use, thus suggesting that such a combination of analytical tools is uniquely informative and promising for analyzing associations relating to emotional dynamics, especially within a longitudinal context. Future efforts may benefit from applying an expanded sentiment reference dictionary or machine learning–based sentiment analysis approaches with the inclusion of *n*-grams to more fully test the practical application and theoretical capabilities of the circumplex model of affect on text-based data. Furthermore, in analyzing news headline content, the present results implicated arousal as the most informative and statistically significant circumplex dimension. Future studies of news content not necessarily limited to COVID-19 or mental health may therefore find it informative to incorporate and focus attention on the arousal-based qualities of word use.

Acknowledgments

This work was funded by an institutional grant from the National Institute on Drug Abuse (NIDA-5P30DA02992610).

Authors' Contributions

All authors made significant contributions to the manuscript. The following reflects individual contributions per the nomenclature of the Contributor Roles Taxonomy (CRediT): DL was responsible for conceptualization, methodology, software, formal analysis, writing—original draft preparation, writing—review and editing, and visualization. JAG was responsible for writing—original draft preparation and writing—review and editing. GDP took care of formal analysis, writing—original draft preparation, writing—review and editing, and visualization. ZW was responsible for methodology and software analysis. NCJ was responsible for formal analysis and writing—review and editing.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table of news outlets and associated details.
[DOC File , 210 KB - jmir_v24i1e32731_app1.doc]

Multimedia Appendix 2
Table of Google Trends comparator search terms by date.
[DOC File , 186 KB - jmir_v24i1e32731_app2.doc]

Multimedia Appendix 3
Table of additional stopwords used in news story title preprocessing.
[DOC File , 40 KB - jmir_v24i1e32731_app3.doc]

Multimedia Appendix 4
The raw statistical results of state-specific affect mean, affect variance, and affect RMSSD across the five temporal windows (three pandemic phases) of the study. Affect is quantified in three dimensions: (i) valence, (ii) arousal, and (iii) dominance.
[XLS File (Microsoft Excel File), 47 KB - jmir_v24i1e32731_app4.xls]

Multimedia Appendix 5
The tabular, long-format data used in negative binomial mixed-modeling to model the association between the affective dynamics of state-specific COVID news headlines (via the affect circumplex) on same-day online mental health search term activity.
[XLS File (Microsoft Excel File), 6030 KB - jmir_v24i1e32731_app5.xls]

Multimedia Appendix 6
Choropleths of state-based random effects from negative binomial mixed models. The conditional modes for each state are plotted for each of the six models. Darker blue coloration denotes increasingly positive associations with the outcome, while darker red coloration denotes increasingly negative associations with the outcome.
[PNG File , 143 KB - jmir_v24i1e32731_app6.png]

References

1. de Hoog N, Verboon P. Is the news making us unhappy? The influence of daily news exposure on emotional states. Br J Psychol 2019 Mar 21;111(2):157-173. [doi: 10.1111/bjop.12389] [Medline: 30900253]

2. McIntyre KE, Gibson R. Positive News Makes Readers Feel Good: A “Silver-Lining” Approach to Negative News Can Attract Audiences. Southern Communication Journal 2016 Jun 07;81(5):304-315. [doi: 10.1080/1041794x.2016.1171892]
3. Shine K, Rogers S. Public perceptions of the tone and emotional impact of educational news coverage. PsyArXiv 2020 Nov 18:1-15. [doi: 10.31234/osf.io/4s6gw]
4. Yadamsuren B, Heinström J. Emotional reactions to incidental exposure to online news. Information Research 2011 Aug 17;16(3):1-20 [FREE Full text]
5. Kramer ADI, Guillory JE, Hancock JT. Experimental evidence of massive-scale emotional contagion through social networks. Proc Natl Acad Sci USA 2014 Jun 17;111(24):8788-8790 [FREE Full text] [doi: 10.1073/pnas.1320040111] [Medline: 24889601]
6. Rieis J, Benevenuto F, Olmo P, Prates R, Kwak H, An J. Breaking the news: First impressions matter on online news. In: Proceedings of the International AAAI Conference on Web and Social Media. 2015 Presented at: Ninth International AAAI Conference on Web and Social Media; May 26-29, 2015; Oxford, UK URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14619
7. Bostan L, Kim E, Klinger R. GoodNewsEveryone: A corpus of news headlines annotated with emotions, semantic roles, and reader perception. In: Conference on Language Resources and Evaluation. 2020 Mar 03 Presented at: Proceedings of the 12th Conference on Language Resources and Evaluation; May 2020; Marseille, France p. 1554-1566 URL: http://arxiv.org/abs/1912.03184
8. Fletcher R, Nielsen RK. Are people incidentally exposed to news on social media? A comparative analysis. New Media & Society 2018;20(7):2450-2468. [doi: 10.1177/1461444817724170]
9. Zhaoyang R, Scott S, Smyth J, Kang J, Sliwinski M. Emotional responses to stressors in everyday life predict long-term trajectories of depressive symptoms. Annals of Behavioral Medicine 2020 May 25;54(6):402-412. [doi: 10.1093/abm/kaz057]
10. Deshmukh R, Kirange D. Emotion classification of news headlines using SVM. Asian Journal of Computer Science and Information Technology 2012;2(5):104-106.
11. Bakshy E, Messing S, Adamic LA. Exposure to ideologically diverse news and opinion on Facebook. Science 2015 Jun 05;348(6239):1130-1132. [doi: 10.1126/science.aaa1160] [Medline: 25953820]
12. Ni MY, Yang L, Leung CMC, Li N, Yao XI, Wang Y, et al. Mental health, risk factors, and social media use during the COVID-19 epidemic and cordon sanitaire among the community and health professionals in Wuhan, China: Cross-sectional survey. JMIR Ment Health 2020 May 12;7(5):e19009. [doi: 10.2196/19009]
13. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: A web-based cross-sectional survey. Psychiatry Research 2020 Apr 12;288:1-6. [doi: 10.1016/j.psychres.2020.112954]
14. Kecojevic A, Basch CH, Sullivan M, Davi NK. The impact of the COVID-19 epidemic on mental health of undergraduate students in New Jersey, cross-sectional study. PLoS One 2020 Sep 30;15(9):e0239696. [doi: 10.1371/journal.pone.0239696] [Medline: 32997683]
15. Huckins JF, daSilva AW, Wang W, Hedlund E, Rogers C, Nepal SK, et al. Mental health and behavior of college students during the early phases of the COVID-19 pandemic: Longitudinal smartphone and ecological momentary assessment study. J Med Internet Res 2020 Jun 17;22(6):e20185. [doi: 10.2196/20185]
16. Aslam F, Awan TM, Syed JH, Kashif A, Parveen M. Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. Humanit Soc Sci Commun 2020 Jul 08;7(1):1-9. [doi: 10.1057/s41599-020-0523-3]
17. Pretorius C, Chambers D, Coyle D. Young people's online help-seeking and mental health difficulties: Systematic narrative review. J Med Internet Res 2019 Nov 19;21(11):e13873 [FREE Full text] [doi: 10.2196/13873] [Medline: 31742562]
18. Gunnell D, Derges J, Chang S, Biddle L. Searching for Suicide Methods. Crisis 2015 Oct 27;36(5):325-331. [doi: 10.1027/0227-5910/a000326]
19. Parker J, Cuthbertson C, Loveridge S, Skidmore M, Dyar W. Forecasting state-level premature deaths from alcohol, drugs, and suicides using Google Trends data. J Affect Disord 2017 Dec 15;213:9-15. [doi: 10.1016/j.jad.2016.10.038] [Medline: 28171770]
20. Soreni N, Cameron DH, Streiner DL, Rowa K, McCabe RE. Seasonality patterns of internet searches on mental health: Exploratory infodemiology study. JMIR Ment Health 2019 Apr 24;6(4):e12974. [doi: 10.2196/12974]
21. Jacobson NC, Lekkas D, Price G, Heinz MV, Song M, O’Malley AJ, et al. Flattening the mental health curve: COVID-19 stay-at-home orders are associated with alterations in mental health search behavior in the United States. JMIR Ment Health 2020 Jun 1;7(6):e19347. [doi: 10.2196/19347]
22. Medhat W, Hassan A, Korashy H. Sentiment analysis algorithms and applications: A survey. Ain Shams Engineering Journal 2014 May 27;5(4):1093-1113. [doi: 10.1016/j.asej.2014.04.011]
23. Sánchez-Rada JF, Iglesias CA. Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. Information Fusion 2019 May 13;52:344-356. [doi: 10.1016/j.inffus.2019.05.003]
24. Zou H, Tang X, Xie B, Liu B. Sentiment classification using machine learning techniques with syntax features. In: International Conference on Computational Science and Computational Intelligence (CSCI).: IEEE; 2015 Dec 09 Presented at: 2015 International Conference on Computational Science and Computational Intelligence (CSCI); December 7-9, 2015; Las Vegas, NV USA. [doi: 10.1109/csci.2015.44]

25. Geethapriya A, Valli S. An enhanced approach to map domain-specific words in cross-domain sentiment analysis. Inf Syst Front 2021 Jan 05;23(3):791-805. [doi: 10.1007/s10796-020-10094-5]
26. Appel O, Chiclana F, Carter J, Fujita H. A hybrid approach to the sentiment analysis problem at the sentence level. Knowledge-Based Systems 2016 May 20;108:110-124. [doi: 10.1016/j.knosys.2016.05.040]
27. Chakraborty K, Bhatia S, Bhattacharyya S, Platos J, Bag R, Hassanien AE. Sentiment analysis of COVID-19 tweets by deep learning classifiers—A study to show how popularity is affecting accuracy in social media. Applied Soft Computing Journal 2020 Sep 28;97:1-14. [doi: 10.1016/j.asoc.2020.106754]
28. Chintalapudi N, Battineni G, Canio MD, Sagaro GG, Amenta F. Text mining with sentiment analysis on seafarers' medical documents. International Journal of Information Management Data Insights 2020 Dec 28;1(1):1-9. [doi: 10.1016/j.jjimei.2020.100005]
29. Bakker I, van der Voordt T, Vink P, de Boon J. Pleasure, arousal, dominance: Mehrabian and Russell revisited. Curr Psychol 2014 Jun 11;33(3):405-421. [doi: 10.1007/s12144-014-9219-4]
30. Munezero M, Montero CS, Sutinen E, Pajunen J. Are they different? Affect, feeling, emotion, sentiment, and opinion detection in text. IEEE Trans. Affective Comput 2014 Apr 1;5(2):101-111. [doi: 10.1109/TAFFC.2014.2317187]
31. Posner J, Russell JA, Peterson BS. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. Dev Psychopathol 2005 Nov 01;17(3):715-734 [FREE Full text] [doi: 10.1017/S0954579405050340] [Medline: 16262989]
32. Yik M, Russell JA, Steiger JH. A 12-point circumplex structure of core affect. Emotion 2011 Aug;11(4):705-731. [doi: 10.1037/a0023980]
33. Huang S, Zhou L, Liu Z, Ni S, He J. Empirical research on a fuzzy model of music emotion classification based on pleasure-arousal model. In: Chinese Control Conference (CCC). 2018 37th Chinese Control Conference (CCC) Internet Wuhan: IEEE; 2018 Jul Presented at: 37th Chinese Control Conference (CCC); July 25-27, 2018; Wuhan, China p. 3239-3244 URL: https://ieeexplore.ieee.org/document/8483910 [doi: 10.23919/chicc.2018.8483910]
34. Naskar D, Mokaddem S, Rebollo M, Onaindia E. Sentiment analysis in social networks through topic modeling. In: Proceedings of the International Conference on Language Resources and Evaluation (LREC).: European Language Resources Association (ELRA); 2016 May Presented at: Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16); May 23-28, 2016; Portorož, Slovenia p. 46-53 URL: https://www.aclweb.org/anthology/L16-1008
35. Preotiuc-Pietro D, Schwartz H, Park G, Eichstaedt J, Kern M, Ungar L, et al. Modelling valence and arousal in Facebook posts. In: Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis. Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis Internet San Diego, California: Association for Computational Linguistics; 2016 Jun Presented at: Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis; June 12-17, 2016; San Diego, CA USA p. 9-15. [doi: 10.18653/v1/w16-0404]
36. Russell JJ, Moskowitz DS, Zuroff DC, Sookman D, Paris J. Stability and variability of affective experience and interpersonal behavior in borderline personality disorder. Journal of Abnormal Psychology 2007 Aug;116(3):578-588. [doi: 10.1037/0021-843x.116.3.578]
37. Chester DS, Clark MA, DeWall CN. The flux, pulse, and spin of aggression-related affect. Emotion 2021 Apr;21(3):513-525. [doi: 10.1037/emo0000730]
38. WHO Coronavirus (COVID-19) Dashboard. World Health Organization. 2021. URL: https://covid19.who.int [accessed 2021-06-28]
39. Newspapers in the United States. 4 International Media & Newspapers. 2021. URL: https://www.4imn.com/us/ [accessed 2021-06-13]
40. Top Media Outlets. Agility PR Solutions. 2015. URL: https://www.agilitypr.com/resources/top-media-outlets/ [accessed 2021-06-13]
41. Media Cloud. 2021. URL: https://mediacloud.org [accessed 2021-06-13]
42. Valiukas L. Media Cloud API. Media Cloud. 2021. URL: https://github.com/mediacloud/backend [accessed 2021-06-13]
43. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (5th ed.). Washington, DC USA: American Psychiatric Association Publishing; 2013.
44. Fernandez KC, Fisher AJ, Chi C. Development and initial implementation of the dynamic assessment treatment algorithm (DATA). PLoS One 2017 Jun 27;12(6):1-16 [FREE Full text] [doi: 10.1371/journal.pone.0178806] [Medline: 28654667]
45. Loper E, Bird S. NLTK: The natural language toolkit. In: Proceedings of the ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics.: Association for Computational Linguistics; 2002 Presented at: The ACL-02 Workshop on Effective Tools and Methodologies for Teaching Natural Language Processing and Computational Linguistics; 2002; Philadelphia, PA USA p. 63-70. [doi: 10.3115/1118108.1118117]
46. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas. Behav Res 2013 Feb 13;45(4):1191-1207. [doi: 10.3758/s13428-012-0314-x]
47. von Neumann J, Kent RH, Bellinson HR, Hart BI. The mean square successive difference. Ann. Math. Statist 1941 Jun;12(2):153-162. [doi: 10.1214/aoms/1177731746]

48. Brooks M, Kristensen K, Benthem K, Magnusson A, Berg C, Nielsen A, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. The R Journal 2017 Dec;9(2):378-400. [doi: 10.32614/rj-2017-066]
49. Windmeijer FA. Goodness-of-fit measures in binary choice models. Econometric Reviews 1995 Jan;14(1):101-116. [doi: 10.1080/07474939508800306]
50. Lüdecke D. sjPlot: Data visualization for statistics in social science (ver. 2.8). The Comprehensive R Archive Network (CRAN) 2021:1-106 [FREE Full text]
51. Grant A. There's a name for the blah you're feeling: It's called languishing. The New York Times. 2021 Apr 19. URL: https://www.nytimes.com/2021/04/19/well/mind/covid-mental-health-languishing.html [accessed 2021-08-05]
52. Maier SF, Seligman ME. Learned helplessness: Theory and evidence. Journal of Experimental Psychology: General 1976;105(1):3-46. [doi: 10.1037/0096-3445.105.1.3]
53. Hutto C, Gilbert E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the International AAAI Conference on Web and Social Media.: PKP Publishing Services Network; 2014 May 16 Presented at: Proceedings of the Eighth International AAAI Conference on Web and Social Media; June 1-4, 2014; Ann Arbor, MI USA p. 216-225 URL: https://ojs.aaai.org/index.php/ICWSM/article/view/14550
54. Dey A, Jenamani M, Thakkar JJ. Senti-N-Gram : An n-gram lexicon for sentiment analysis. Expert Systems with Applications 2018 Aug;103:92-105. [doi: 10.1016/j.eswa.2018.03.004]

Abbreviations

API: application programming interface
NLP: natural language processing
RMSSD: root mean square of successive differences
VA: valence–arousal
VAD: valence–arousal–dominance

Edited by C Basch; submitted 08.08.21; peer-reviewed by Y Zhang, P Verboon; comments to author 05.11.21; revised version received 29.11.21; accepted 30.11.21; published 27.01.22.

Please cite as:
Lekkas D, Gyorda JA, Price GD, Wortzman Z, Jacobson NC
Using the COVID-19 Pandemic to Assess the Influence of News Affect on Online Mental Health-Related Search Behavior Across the United States: Integrated Sentiment Analysis and the Circumplex Model of Affect
J Med Internet Res 2022;24(1):e32731
URL: https://www.jmir.org/2022/1/e32731
doi:10.2196/32731
PMID:34932494

Original Paper

Understanding Health Empowerment From the Perspective of Information Processing: Questionnaire Study

Fei Jiang[1*], PhD; Yongmei Liu[1*], PhD; Junhua Hu[1*], PhD; Xiaohong Chen[2*], PhD

[1]Business School of Central South University, Changsha, China

[2]Hunan University of Technology and Business, Changsha, China

[*]all authors contributed equally

Corresponding Author:
Yongmei Liu, PhD
Business School of Central South University
No 932, Lushan South Road
Yuelu District
Changsha, 410083
China
Phone: 86 13974834821
Email: liuyongmeicn@163.com

Abstract

Background: Massive, easily accessible online health information empowers users to cope with health problems better. Most patients search for relevant online health information before seeing a doctor to alleviate information asymmetry. However, the mechanism of how online health information affects health empowerment is still unclear.

Objective: To study how online health information processing affects health empowerment.

Methods: We conducted a cross-sectional questionnaire study that included 343 samples from participants who had searched online health information before the consultation. Respondents' perceptions of online information cues, benefits, health literacy, and health empowerment were assessed.

Results: Perceived argument quality and perceived source credibility have significant and positive effects on perceived information benefits, but only perceived argument quality has a significant effect on perceived decision-making benefits. Two types of perceived benefits, in turn, affect health empowerment. The effects of perceived argument quality on perceived informational benefits and perceived decision-making benefits on health empowerment are significantly stronger for the high health literacy group than the low health literacy group (t_{269}=7.156, P<.001; t_{269}=23.240, P<.001). While, the effects of perceived source credibility on perceived informational benefits and perceived informational benefits on health empowerment are significantly weaker for the high health literacy group than the low health literacy group (t_{269}=–10.497, P<.001; t_{269}=–6.344, P<.001). The effect of perceived argument quality on perceived informational benefits shows no significant difference between high and low health literacy groups.

Conclusions: In the context of online health information, perceived information benefits and perceived decision-making benefits are the antecedents of health empowerment, which in turn will be affected by perceived argument quality and perceived source credibility. Health literacy plays a moderating role in the relationship of some variables. To maximize health empowerment, online health information providers should strengthen information quality and provide differentiated information services based on users' health literacy.

(J Med Internet Res 2022;24(1):e27178) doi:10.2196/27178

KEYWORDS

online health information; perceived argument quality; perceived source credibility; health literacy; health empowerment; information seeking

Introduction

Background

Health empowerment is a cornerstone of a patient-centered approach to healthcare. Empowerment allows patients to take the initiative in making decisions about their own health care and quality of life, rather than passively complying with decisions made by others [1,2]. Previous literature hailed health empowerment as a new paradigm for health management and nursing practice [3,4]. As such, how to promote individuals' health empowerment has become a common concern of scholars and health care professionals.

The rise of e-health services has brought new opportunities for promoting health empowerment. Various forms of electronic health services (eg, health information portals, online health communities, consultation platforms, etc) provide the public with abundant and easily accessible health information. Patients can obtain information about the symptoms of the disease, conventional treatment methods, and the treatment experience of others. With that health information, patients can become informed before doctors' visits and participate in health decision-making during the consultation process to enhance their sense of control. And an increasing number of people now obtain health information online. The China Internet Network Information Center pointed out that more than 276 million users in China utilize internet medical services, accounting for 29.4% of all internet users [5]. Online health information is changing the traditional way of the doctor visit, in which patients passively follow doctor's decisions.

Research on health empowerment in the context of eHealth services has become an important research stream. Some scholars explored the logic or dimensions of empowerment in the context of eHealth services [6,7]. Other scholars focused on the promotion of health empowerment by the benefits or functions of online health services. Electronic health records make users more informed and in a favorable position in the medical market [8]. Online health communities can provide users with various social support to promote empowerment [9-12]. Berkel et al [13] showed patients discussing drug use information on online message boards can promote patient empowerment, and the most common empowerment process is providing information and sharing personal experiences. Nelson et al [14] extracted the six system elements of wearable devices and pointed out that they can promote the user's health empowerment and commitment to health goals. In addition, some scholars have pointed out that web-based interventions can promote patient empowerment [15-17].

Previous research provided us with valuable knowledge for understanding health empowerment. Undoubtedly, obtaining health information from online resources to reduce information asymmetry is an indispensable part of patient empowerment [18,19]. However, perhaps we should pay more attention to the mechanism of information processing on health empowerment currently. Accessing online health information is easy nowadays due to high internet penetration, available devices, available information, but once information has been accessed, processing information is a crucial next step. In the context of processing online health information, individuals with different health literacy may face different situations. Health literacy measures the ability to acquire, process, and understand basic health information and the ability to use health information to make healthy decisions [20]. The usefulness of online health information largely depends on the recipients' health literacy [21]. Even in the face of the same online health information, the receivers with different health literacy will have different perceptions and health empowerment. In order to promote health empowerment more efficiently, it is necessary to focus further on information processing, explore how information recipients benefit from online health information, and ultimately promote health empowerment. Therefore, this study focuses on two issues: (1) How do users' processing of health information contribute to health empowerment? (2) How does health literacy affect an individual's health empowerment process?

Overall, we assume that online health information can promote health empowerment during the consultation process, which is the result of the interaction information factors and the health literacy of the recipient. As a popular health resource, online health information can support patients with the ability to participate in the consultation process. Therefore, it is necessary to explore the process by which patients analyze online information and identify the mechanisms by which they contribute to health empowerment. To address this question, based on the elaboration likelihood model (ELM), we conceptualized perceived argument quality, perceived source credibility, and health literacy into online health information processing scenarios and explored their impact on health benefits and health empowerment.

Literature Review

Health Empowerment

Empowerment theory has been explored by a rich body of research in social work, mainly as it relates to self-esteem, self-worth, self-confidence, and wellness [22,23]. Health empowerment is a further development of empowerment theory in the medical field. As part of a patient-centered philosophy, health empowerment emphasizes that individuals are responsible for their own health [24]. Health empowerment focuses on keeping individuals informed, encouraging active patient participation in decision-making [25,26], and working toward individual self-efficacy with regard to health matters [27,28].

Although health empowerment has been one of the core concepts in health promotion research, there is still no unified definition. Past researches have mainly defined health empowerment from three perspectives: process, emergent state, and active behavior [29]. From a process perspective, health empowerment is defined as the process leading to personal transformation, through which the individual's ability to cope with health problems is developed [30,31]. In general, the implementation of the empowerment process requires the support of external resources. From an emergent state perspective, health empowerment represents the individual's health skills and psychological cognition, such as the health knowledge, skills, attitudes, and self-awareness with which people can make better health decisions [32]. This definition highlights motivation and ability and assesses an individual's state of being empowered.

From an active behavior perspective, health empowerment is interpreted as the actual behavior change after possession, ability, and motivation.

Due to different definitions and research contexts, previous studies have used multidimensional or single-dimensional assessments of health empowerment. Ouschan et al [33] proposed that empowerment in the context of medical consultation includes three dimensions: patient control, patient participation, and doctor support. Prigge et al [34] understand empowerment as the behaviors that meet the inherent needs of autonomy and competence, including three dimensions: information search, knowledge development, and decision-making participation. From the perspective of the internal motivation process, Londono and Schulz [35] evaluate health empowerment in four dimensions: meaning, competence, self-determination, and impact. There are also some studies that assess health empowerment from a single dimension [36,37].

This study considers health empowerment from the perspective of the state of being empowered. Accessible online health information eases the information asymmetry between doctors and patients to a certain extent. The patient is no longer in a completely passive position but can actively participate in health activities. This undoubtedly allows patients to advocate for themselves and increase their sense of control. We define health empowerment as one's belief that they have a significant influence over health outcomes, including the ability to address personal health issues and feel in control over factors that can impact health outcomes.

Elaboration Likelihood Model

The elaboration likelihood model (ELM) explains how two types of information persuasion paths affect individuals' attitude changes, perceptions, and behaviors [38]. The ELM has been used for many information systems literature as the theoretical basis for researching information adoption [39,40], online physician selection [41], and information technology adoption contexts [42,43]. The model postulates that external information can lead to attitude changes by two means: the central route and the peripheral route. The two routes distinguish one another in terms of the level of cognitive effort involved in processing information [38]. For the central route, persuasion results from careful consideration of the arguments regarding the core issues presented by the information. The recipients exert a high degree of cognitive effort. For the peripheral route, persuasion does not come from the information itself but from nonissue-related concerns, and the recipients devote less cognitive effort to the process [38,44]. The influence of each of the two routes can cause attitude changes and consequent behavior changes, but the changes caused by the central route are usually more stable and long-lasting than those caused by the peripheral route [45,46].

In addition, the ELM generally approaches elaboration likelihood from two influencing dimensions: ability and motivation [45]. If information recipients view a given message as being important or have a greater belief that they are capable of processing the information, they are more likely to invest the needed cognitive effort. In contrast, if recipients view the same message as having little personal relevance, or if they believe that as nonexperts, they have little choice but to depend on peripheral cues, they may be unwilling to spend much time and effort to scrutinize the information content [40]. Hence, ability and motivation are generally considered to moderate the relationship between two types of routes and perception changes [47]. Typically, ELM researchers have operationalized central route processing in terms of perceived argument quality and peripheral route processing in terms of perceived source credibility. Perceived argument quality measures whether the information content provides sufficient reasoning or support to prove the validity of key claims [48], while perceived source credibility measures the reliability and perceived acceptance of the information provider [49].

Research Model and Hypotheses

The Influence of Perceived Informational Benefits and Perceived Decision-making Benefits on Health Empowerment

By providing online health information and educational opportunities, information and communication technology (ICT) can empower users to deal with health issues and engage in their own health outcomes [50,51]. Assessing consumers' perceived benefits from the use of ICT can enable health professionals and researchers to develop better strategies for using ICT as an empowerment tool to support users in accessing information and managing health. Perceived informational benefits reflect the users' ability to better understand their own health status and treatment options with the support of online information. Perceived decision-making benefits measure the extent to which users can participate effectively in decision-making for their own well-being with the help of knowledge or experiences obtained from the internet [26]. Health empowerment is based on the premise that the individual can obtain relevant medical knowledge and skills [52]. The availability of online health information allows users to acquire the knowledge and skills they need to enhance their self-efficacy. This knowledge allows them to be more confident about participating in treatment decisions by addressing questions to their physicians, sharing feelings, and otherwise being actively involved in their own health care [53].

Online support groups enable patients to learn more about themselves, enhance their social well-being, and thus promote healthy empowerment [11]. Johnston et al [19] explored the impact of participation in online health communities on health empowerment from the perspective of information utility. Their findings showed that online health communities could provide participants with direct benefits such as practical information and social support to further promote their health empowerment.

Since involvement in health consultation and decision-making processes is an important element of health empowerment [54], individual participation in the medical decision-making process will help patients understand medical practices, maximize individual satisfaction, and achieve a better quality of care [55,56]. Health empowerment can be improved by developing individuals' ability to participate actively in the medical decision-making process [57]. With the increasing availability of online health information, individuals can better interact with

their physicians, evaluate services more accurately, and make informed decisions. Therefore, we proposed the following hypotheses:

- **H1:** Perceived informational benefits have a positive impact on health empowerment.
- **H2:** Perceived decision-making benefits have a positive impact on health empowerment.

The Influence of Perceived Argument Quality on Perceived Benefits

Perceived argument quality is reflected in an individual's subjective evaluation of the reasoning that forms the core of presented information. The presentation of information can be strong and convincing or weak and specious. Strong arguments mean that the presented information is reasonable and convincing to the recipient, while weak arguments are doubtful or contradictory [38,58]. In the ELM, perceived argument quality that follows the central path of cognitive processing is an important factor affecting attitudes and decision-making [45,59]. In the information adoption model proposed by Sussman and Siegal [40], perceived argument quality was used as a predictor of the perceived usefulness of the information. Their empirical results showed that as the quality of information arguments increased, the perceived usefulness and adoption intention increased as well. In health information research, the literature has pointed out that perceived argument quality exerts an influence on recipients' attitude changes and is an important index used to evaluate the quality of information [44,60]. The quality of information is directly related to whether information seekers can obtain the information they need and make high-quality decisions [61]. Therefore, perceived argument quality will affect the individual's perceived informational and decision-related benefits. When confronted with the uneven quality of online health information, the recipients can judge the quality and usefulness of the information according to the quality of the arguments to obtain complete, accurate, and validated online health information. In line with the idea that quality information can help people better understand their own health status and perform well in making health decisions, the following hypotheses were proposed:

- **H3a:** The perceived argument quality has a positive impact on the perceived informational benefits.
- **H3b:** The perceived argument quality has a positive impact on the perceived decision-making benefits.

The Influence of Perceived Source Credibility on Perceived Benefits

Perceived source credibility is the evaluation of information from the reliability of information sources. It can be perceived to be credible, acceptable, or untrustworthy by information recipients [62]. A highly credible information source is more persuasive than a less creditable one [49,63]. In the ELM, perceived source credibility as a peripheral route factor affects the attitude change of the information recipient [45,59]. In the context of information adoption, perceived source credibility has been identified as peripheral clues of the given messages that affect information's usefulness [40]. In the context of consumer-to-consumer communication, the credibility of information derived from communication can help consumers evaluate the quality of products, allowing the consumers to make reasonable purchase decisions [64].

In the health information literature, perceived source credibility is an important topic that relates to individual health outcomes and decision-making behavior. Young people's trust in health information is affected by perceived source credibility. The higher credibility of the information source, the more likely the users are to participate in the information activity [65,66]. Ghaddar et al [67] demonstrated that exposure to credible sources of health information could improve individual health literacy. High-quality health information is the basis for individuals to improve their health knowledge and participate in treatment decisions. In addition, source credibility is an important factor used by individuals to evaluate the quality of online health information [68,69]. When faced with uncertain quality online health information, highly credible sources can reduce perceived risk and increase trust in health information [44]. Individuals who obtain more credible health information can better understand their own health conditions and participate more effectively in health decision-making processes, whereas individuals who possess unreliable health information may be led to negative outcomes [70]. Hence, we hypothesized:

- **H4a:** The credibility of sources has a positive effect on the perceived informational benefits.
- **H4b:** The credibility of sources has a positive effect on the perceived decision-making benefits.

Moderating Effect of Health Literacy

According to the ELM, the ability to process information can affect the level of elaboration likelihood [45]. In our work, health literacy was identified to measure ones' ability to process health information gained online. Health literacy measures the ability to acquire, process, and understand basic health information, as well as the ability to use health information to make healthy decisions [20]. Previous research demonstrated that health literacy correlates with individuals' health information acquisition and use behaviors. Individuals with adequate health literacy are more inclined to access health information through multiple channels, such as the internet, rather than relying solely on medical personnel [71,72]. Ghaddar et al [67] also pointed out that health literacy positively affects the self-efficacy and motivation of individuals to gather the information needed from online sources. In addition, health literacy determines the effective response of information recipients to health information to a certain extent. Individuals with limited health literacy can obtain more information as needed if the information providers reduce the cognitive requirements for understanding online information [21].

In the process of health information analysis, individuals with adequate health literacy have a greater ability to analyze the arguments presented as part of health information. For individuals whose attitudes or perceptions change based on central route processing, the information influence occurs under conditions of high-end elaboration (ie, content-oriented reasoning). In contrast, for individuals with limited health literacy, information processing is more about evaluating factors other than content, so peripheral cues play a more critical role

in processing. These expectations led us to state the following hypotheses:

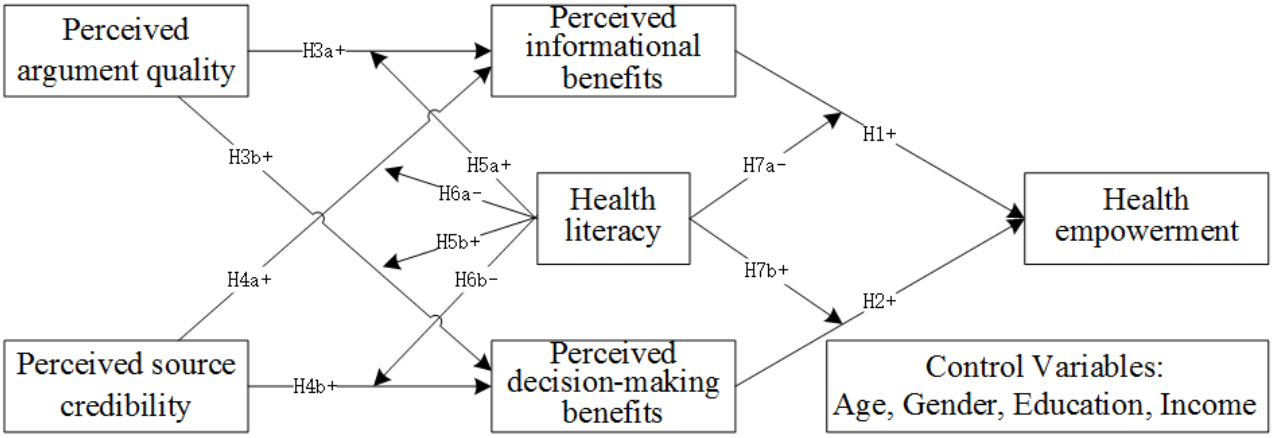
- **H5a:** For users with high health literacy, perceived argument quality has a stronger impact on perceived informational benefits than that of users with low health literacy.
- **H5b:** For users with high health literacy, perceived argument quality has a stronger impact on perceived decision-making benefits than that of users with low health literacy.
- **H6a:** For users with low health literacy, perceived source credibility has a stronger impact on perceived informational benefits than that of users with high health literacy.
- **H6b:** For users with low health literacy, perceived source credibility has a stronger impact on perceived decision-making benefits than that of users with high health literacy.

We assume that the perceived informational benefits and decision-making benefits all contribute to empowerment. However, the two kinds of benefits have different requirements for patients' health literacy. Informational benefits are the prerequisite for decision-making benefits. Sufficient information can improve the quality of decision-making and reduce risks [73]. Meanwhile, decision-making requires patients to devote more cognitive costs. Because decision-making means that patients have to make some trade-offs, such as choosing a suitable therapy among multiple treatment options[74], these require patients to organize and process health information at a deeper level, which means that adequate health literacy is needed. Individuals with high health literacy tend to participate in health decisions [74,75]. Therefore, patients have varying degrees of demands or expectations for informational benefits and decision-making benefits in promoting empowerment. For groups with high health literacy, the appeal of rights and interests in the consultation is not limited to getting more information but also encourages participation in the decision-making process. For groups with low health literacy, their information processing capacity is insufficient to form effective health decisions; their appeals focus on obtaining informational benefits. Therefore, we get the following hypotheses:

- **H7a:** For users with high health literacy, perceived informational benefits have a weaker impact on health empowerment than that of users with low health literacy.
- **H7b:** For users with high health literacy, perceived decision-making benefits have a stronger impact on health empowerment than that of users with low health literacy.

A summary of the conceptual research model is depicted in Figure 1.

Figure 1. Research model.



Methods

Measurement Development

To test our hypotheses, we administered a self-reported questionnaire to collect data. The questionnaire consisted of two parts: one was designed to investigate the demographic characteristics of the participants, and the other focused on the measurement of the constructs. The research model contained a total of 6 constructs. The measurement scales were developed by drawing on prior literature, and some items were fine-tuned according to the background of this study. We adapted the work of Hur et al [76] to measure perceived argument quality (eg, "the health information provided online is informative" and "the health information provided online is persuasive") and the work of Sussman and Siegal [40] to measure perceived source credibility (eg, "the provider of online health information is knowledgeable" and "the provider is an expert on the message topic"). The measurement of perceived informational benefits (eg, "by searching for online health information, I feel better informed as a patient" and "by searching for online health information, I understand my illness better") was adapted from the paper by van Uden-Kraan et al [11], and the measurement of perceived decision-making benefits (eg, "online health information is helpful to decide what questions to ask during doctor appointments" and "online health information is helpful to decide on treatment choices and make decisions") was adapted from Seçkin [26]. The determination of health literacy (eg, "I know how to use the internet to answer my questions about health" and "I have the skills I need to evaluate the health resources and information I find on the internet") was adapted from the eHealth Literacy Scale [77]. The items used to measure health empowerment (eg, "I feel more in control of my health" and "I know what to do to take care of my health problem")

were adapted from Bann et al [36]. There are two reasons for the choice. First, the content captured by the empowerment in this study is similar to Bann et al [36], that is, enablement and the sense of control. Although the two research contexts are different, both explore the improvement of patients' ability and a sense of control with the support of external convenience. Second, the scale has been used in many studies [78-80], and it has been proven to have good reliability and validity. For all constructs, measures were designed using 5-point Likert scales from 1 (“strongly disagree”) to 5 (“strongly agree”).

Since the respondents are Chinese, we need to translate all the items from English into Chinese. All measures were back-translated by another translator who did not know the background of the study to ensure the accuracy of the translation. The two English versions were compared, and potential semantic discrepancies were examined to ensure that the Chinese scales reflected the meaning of all measures accurately. Then 10 postgraduates with experience seeking online health information were invited to participate in a pretest of the scales. Based on their feedback, any ambiguous expressions were amended. The measured constructs and their sources are shown in Multimedia Appendix 1.

Survey Administration

We collected data through a questionnaire service website [81]. The survey participants were required to have had experience seeking online health information before they consulted a doctor during the most recent 6-month period to ensure that participants had an accurate understanding of each measurement item. In addition, subjects were asked to evaluate each measurement item based on their last used experience. A total of 371 questionnaires were collected within 2 weeks. We deleted 28 uncompleted or invalid questionnaires, leaving a total of 343 valid responses. The response rate was 92.7%.

Among the valid questionnaires, 47.2% (162/343) were from males, and 52.8% (181/343) were from females. Further, 85.1% (292/343) of respondents’ ages ranged from 18-35 years, implying that the majority of online health information users tend to be younger. In terms of education, 88.9% (305/343) of the respondents had a college degree or above. The majority (234/343, 68.2%) of the respondents had a monthly disposable income in the range of 3000-8999 Chinese Yuan (approximately US $469-1406). As to their occupations, business employees accounted for the largest proportion of participants, reaching 47.5% (163/343). The most popular way to access information was through a health information portal, accounting for 61.8% (212/343) of the respondents, followed by a health consulting platform, accounting for 22.7% (78/343). On average, 63.3% (217/343) of the respondents used online health information sources between 1 and 3 times weekly, and 22.4% (77/343) of the subjects used these sources 4 to 5 times weekly. The specific demographic information of the target samples is shown in Table 1.

Table 1. Demographic information of respondents (N=343).

Characteristics	Participants, n (%)
Gender	
Male	162(47.2)
Female	181(52.8)
Age, years	
18-25	89(25.9)
26-35	203(59.2)
36-45	45(13.2)
46 and above	6(1.7)
Education	
High school or below	38(11.1)
Associate degree	101(29.4)
College degree	176(51.3)
Master degree or above	28(8.2)
Income, Chinese Yuan[a]/month	
Under 3000	83(24.2)
3000—5999	148(43.1)
6000—8999	86(25.1)
9000—11,999	16(4.7)
12,000 and above	10(2.9)
Occupation	
Student	41(12)
Business employees	163(47.5)
Government and public institutions	39(11.4)
Self-employed persons	45(13.1)
Other	55(16)
Weekly usage frequency (times)	
1-3	217(63.3)
4-5	77(22.4)
6-7	25(7.3)
7 and above	24(7)
Information channel	
Health information portal	212(61.8)
Online patient community	36(10.5)
Health consultation platform	78(22.7)
Blog or video	8(2.3)
Other	9(2.6)

[a]A currency exchange rate of ¥1 = US $0.16 is applicable.

Results

Overview

We used variance-based partial least squares structural equation modeling (PLS-SEM) for data analysis. We chose the PLS-SEM method for the following reasons. First, the PLS-SEM method does not require multivariate normal distribution data [82]. We performed Kolmogorov-Smirnov test (K-S test) to examine the distribution of sample data. And we found that the significance level of all items is less than 0.05. Therefore, the null hypothesis is rejected, meaning the data is nonnormally distributed. Second,

PLS-SEM is suitable for exploratory research because it aims at theoretical development rather than the confirmation of the established theory [82,83]. Finally, the PLS-SEM method has fewer restrictions on the sample size [84]. Compared with other methods, it can obtain greater statistical power with nonlarge sample size. We first examined the measurement model and then the structural model.

Reliability and Validity Analysis

In this study, we used the confirmatory factor analysis process to test the measurement model. As shown in Table 2, Cronbach's α of all constructs is between 0.717 and 0.895. And the composite reliability of each construct is between 0.823 and 0.916. These are above the recommended value of 0.7, which means that the measurement model has good reliability [85]. To assess convergent validity, we measured the standard loading of each item as well as the average variance extracted (AVE) for each construct. The results showed that the items' loadings range from 0.678 to 0.833. Among them, two items' loadings (PAQ4 and PDB1) are less than 0.7 but still much larger than the cutoff value of 0.6 [86]. Also, the AVE of each construct surpasses 0.5. These results imply that the measurement model has good convergence validity [85].

Furthermore, as shown in Table 3, the square root of the AVE of each construct is larger than its correlation coefficients with other constructs, which means the discriminant validity of the measurement model is confirmed [85].

Table 2. Results of confirmatory factor analysis.

Construct and item	Loading	Cronbach's α	Composite reliability	AVE[a]
Perceived argument quality (PAQ)		0.764	0.850	0.587
PAQ1	0.750			
PAQ2	0.827			
PAQ3	0.784			
PAQ4	0.695			
Perceived source credibility (PSC)		0.802	0.870	0.627
PSC1	0.785			
PSC2	0.781			
PSC3	0.793			
PSC4	0.808			
Perceived informational benefits (PIB)		0.756	0.845	0.578
PIB1	0.787			
PIB2	0.747			
PIB3	0.708			
PIB4	0.797			
Perception decision-making benefits (PDB)		0.717	0.823	0.539
PDB1	0.678			
PDB2	0.762			
PDB3	0.724			
PDB4	0.769			
Health empowerment (EM)		0.786	0.854	0.539
EM1	0.736			
EM2	0.775			
EM3	0.702			
EM4	0.736			
EM5	0.720			
Health literacy (HL)		0.895	0.916	0.578
HL1	0.773			
HL2	0.833			
HL3	0.787			
HL4	0.712			
HL5	0.715			
HL6	0.758			
HL7	0.750			
HL8	0.747			

[a]AVE: average variance extracted.

Table 3. Means, SD, and correlation matrix.

Variable	Mean	SD	PAQ[a]	PSC[b]	PIB[c]	PDB[d]	EM[e]	HL[f]
PAQ	3.910	0.680	**0.766**	—[g]	—	—	—	—
PSC	3.625	0.725	0.624	**0.701**	—	—	—	—
PIB	3.918	0.722	0.634	0.553	**0.760**	—	—	—
PDB	3.812	0.656	0.514	0.409	0.632	**0.734**	—	—
EM	3.676	0.651	0.424	0.410	0.454	0.402	**0.734**	—
HL	3.433	0.799	0.569	0.550	0.593	0.553	0.603	**0.760**

[a]PAQ: perceived argument quality.

[b]PSC: perceived source credibility.

[c]PIB: perceived informational benefits.

[d]PDB: perceived decision-making benefits.

[e]EM: health empowerment.

[f]HL: health literacy.

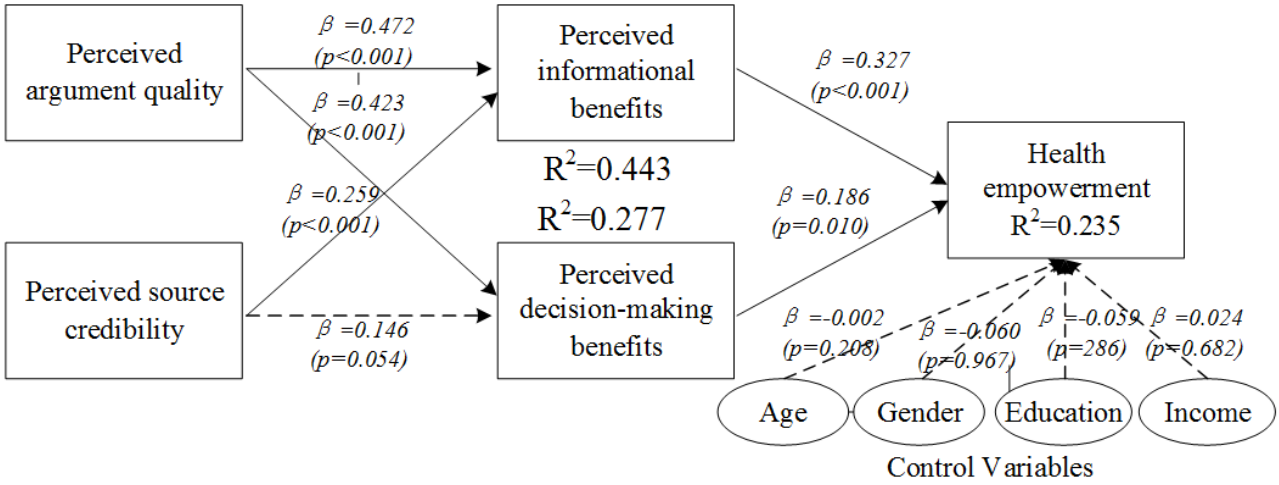
[g]—: The correlation matrix is symmetrical; therefore, only the lower-left corner is displayed.

As our data were collected from single respondents, common method variance (CMV) may threaten the validity of the results. To test such bias, first, we used Harman's single-factor test to assess the 6 constructs in the search model. The results showed that the variance explained by the first factor is 35.4%, which does not exceed 50% [87]. Second, we verified the issue by using the potential factor method from Liang et al [88]. We introduced a common method factor into the PLS-SEM model, which contains all the constructs' indicators. Then we calculated to what extent the common method factor and the main constructs explain the variance of each indicator, respectively. As shown in Multimedia Appendix 2, the average explain variance based on major constructs and the common method factor is 0.579 and 0.006, respectively, with a larger ratio between them. In summary, CMV should not be a serious concern for this study.

Test of Main Effects

In this paper, we used SmartPLS 3.0 (SmartPLS GmbH) to test the research model. The path coefficients and significance levels of main effects are shown in Figure 2. Perceived informational benefits (β=.327, *P*<.001) and perceived decision-making benefits (β=.186, *P*=.01) exerted positive effects on users' health empowerment, indicating that H1 and H2 were supported. Perceived argument quality had a positive effect on perceived informational benefits (β=.472, *P*<.001) and on perceived decision-making benefits (β=.423, *P*<.001), indicating that H3a and H3b were supported. Perceived source credibility had a positive effect on perceived informational benefits (β=.259, *P*<.001), but the effect on perceived decision-making benefits was not significant. Thus, hypothesis H4a was supported, but H4b was not supported. All control variables had no significant effect on health empowerment. The variances explained by perceived informational benefits, perceived decision-making benefits, and healthy empowerment were 44.3%, 27.7%, and 23.5%, respectively.

Figure 2. PLS Analysis of main effects. PLS: partial least squares.



Test of Moderating Effects

A multigroup comparison method developed by Keil et al [89] was used to test the moderating effect of health literacy. This method tests the moderating effect by evaluating the difference in path coefficients between subgroups and has been used in many studies [90,91]. We first divided the samples into high health literacy and low health literacy groups by the median (3.75), with sample sizes of 174 (≥3.75) and 169 (<3.75). Then we used the data of each subgroup to test the research model and compared the path coefficients.

As shown in Table 4, the path coefficients from perceived argument quality to perceived informational benefits and perceived decision-making benefits to empowerment are significantly stronger for the high health literacy group than the low health literacy group (t_{341}=7.156, P<.001; t_{341}=23.240, P<.001). The path coefficients from perceived source credibility to perceived informational benefits and from perceived informational benefits to empowerment are significantly weaker for the high health literacy group than the low health literacy group (t_{341}=–10.497, P<.001; t_{341}=–6.344, P<.001). Therefore, H5a, H6a, H7a, and H7b were supported. The path coefficients from perceived argument quality to decision-making benefits show no significant difference between the two subgroups. Therefore, H5b was not supported. Although path coefficients perceived source credibility to perceived decision-making benefits show a significant difference, they are not significant in either group.

Table 4. The results of moderating effects.

Paths	High health literacy (n=174)		Low health literacy (n=169)		t_{341} values comparing the two groups
	Coefficient	SE	Coefficient	SE	
PAQ[a]→PIB	0.464	0.099	0.395	0.078	7.156
PAQ→PDB	0.358	0.107	0.341	0.099	1.526
PSC[b]→ PIB	0.181	0.076	0.27	0.081	–10.497
PSC→PDB	0.067	0.122	0.151	0.102	–6.908
PIB[c]→EM[d]	0.190	0.094	0.266	0.126	–6.344
PDB[e]→EM	0.292	0.089	–0.003	0.141	23.240

[a]PAQ: perceived argument quality.

[b]PSC: perceived source credibility.

[c]PIB: perceived informational benefits.

[d]EM: health empowerment.

[e]PDB: perceived decision-making benefits.

Test of Mediating Effects

The bootstrap method was used for the analysis of mediating effects [92]. This method can directly test the indirect effects of independent variables on the dependent variables and does not require the mediating effects to follow normal distribution [93]. Using SmartPLS 3.0, we performed bootstrap with 5000 resamples to obtain a 95% CI for indirect effects and direct effect. According to the results in Table 5, the direct effect of perceived argument quality on health empowerment was not significant. Meanwhile, the indirect effects of perceived argument quality on health empowerment (ie, PAQ→PIB→EM and PAQ→PDB→EM) were significant. It means that perceived informational benefits and perceived decision-making benefits play a fully mediating role between perceived argument quality and health empowerment. The direct effect of perceived source credibility on health empowerment was significant, and the indirect effect of the two variables (ie, PSC→PIB→EM) was also significant. This means that the effect of perceived source credibility on health empowerment was partially mediated by perceived informational benefits. With perceived decision-making benefits as the mediate variable, the indirect effect of PSC on health empowerment was not significant, indicating that perceived decision-making benefits have a nonmediating role in the effect of perceived source credibility on health empowerment.

Table 5. The results of the mediation effect test.

Indirect path	95%CI	Direct path	95%CI	Result
PAQ[a]→PIB[b]→EM[c]	0.007 to 0.173	PAQ→EM	–0.011 to 0.256	Full
PAQ→PDB[d]→EM	0.005 to 0.129	PAQ→EM	–0.011 to 0.256	Full
PSC[e]→PIB→EM	0.004 to 0.097	PSC→EM	0.058 to 0.285	Partial
PSC→PDB→EM	–0.002 to 0.061	PSC→EM	0.058 to 0.285	None

[a]PAQ: perceived argument quality.

[b]PIB: perceived informational benefits.

[c]EM: health empowerment.

[d]PDB: perceived decision-making benefits.

[e]PSC: perceived source credibility.

Discussion

Principal Findings

Based on the ELM model, this paper examined the influencing factors of health empowerment in the context of processing online health information. Our empirical research provided the following results:

First, perceived informational and decision-making benefits are important predictors of users' health empowerment. Perceived informational benefits accrue when individuals become more informed by browsing online health information. This input allows them to have a more objective understanding of their illnesses and health situations, thereby reducing negative emotions, such as anxiety and panic. Perceived decision-making benefits refer to growth in terms of knowledge and skills gained through seeking online health information. This improvement in decision-making capacity allows individuals to participate more effectively in the consultation process and make reasonable suggestions for treatment. The gain of these two kinds of benefits makes users feel empowered.

Second, the results confirm that perceived argument quality, as involved with the central route, has a positive effect on perceived informational and decision-making benefits, while perceived source credibility, which relies on the peripheral route, only has a significant impact on perceived informational benefits. When getting health information from online channels, the strength of the arguments and credibility of the sources reflect the quality of information. They are the guarantee that users can benefit from the information they receive. Both high-quality arguments and credible sources can enhance an individual's acceptance and approval of the information, thus promoting the perceived informational benefits. Individuals need knowledge and skills to make informed health decisions. The online health information presented with high-quality arguments can provide recipients with health knowledge and treatment experience so they can make informed decisions in medical consultations.

Our results show that the credibility of sources has no significant influence on perceived decision-making benefits. One possible explanation is that the basis for supporting individuals' participation in decision-making may come more from the information itself, which develops individuals' knowledge or skills. However, the credibility of information resources as a peripheral cue does not improve knowledge or skill levels and thus cannot support the individual's participation in the decision-making process.

Third, we also found that the effects of the central route and the peripheral route are different in low and high health literacy groups. For individuals with high health literacy, the effect of central route processing (perceived argument quality) on perceived informational benefits is stronger than the influence of processing using peripheral cues (perceived source credibility). Individuals with high health literacy are more likely to exert cognitive effort when assessing the arguments provided by online information. For these individuals, information source credibility is used as a secondary consideration and has a weaker effect on perceived informational benefits. The opposite is true for individuals with low health literacy. For low health literacy groups, their judgments of online health information rely more on the source credibility.

Finally, the study demonstrates that the effects of the two perceived benefits on health empowerment are different between groups with high and low health literacy. The effect of perceived informational benefits on health empowerment is greater in the low health literacy group than in the high health literacy group. The effect of perceived decision-making benefits on health empowerment is significant in the high health literacy group but not in the low-health literacy group. The results show that there is a higher demand for health empowerment for individuals with high health literacy. Merely information benefits are not enough to promote health empowerment but to further obtain perceived decision-making benefits. For low health literacy groups, health empowerment does not derive from participating in decision-making but from getting enough information to reduce information asymmetry.

Theoretical Implications

There are two theoretical contributions of this study. First, the study provided a profound understanding of the mechanism of health information processing on health empowerment. Previous studies highlighted the convenience and positive health outcomes that can be derived from information technology and online health information [9,10,14]. However, fewer studies have focused on how the process of information processing influence users' perceived benefits and health empowerment. To advance this line of research, this study explored how the

two information processing routes affect the individuals' perceived benefits and health empowerment. In this way, our work enriches the existing research on health empowerment promotion.

Second, this study explained the relationship between health literacy and health empowerment from a new perspective that is different from previous literature, which always explores the direct relationship between health literacy and health empowerment [94-96]. This study found that individuals with different health literacy have differences in the processing of online health information. Individuals with higher health literacy tend to focus on the central route, while those with lower health literacy focus on peripheral cues. The findings of this study provide a new perspective for studying the relationship between health literacy and health empowerment in other contexts.

Practical Implications

Based on our theoretical analysis and empirical results, the following practical implications should be noted. First, encouraging patients to search for high-quality online health information is an effective way to promote their health empowerment. The information provider can strengthen information quality management in terms of perceived argument quality and perceived source credibility. Accordingly, to prevent the dissemination of misleading content, online health information providers should establish reasonable evaluation and testing mechanisms. They should strictly scrutinize every piece of health information provided to consumers and ensure that information content is complete, rigorous, sound, and scientific.

Second, information providers should also consider the health literacy of recipients while providing health information. To improve the effectiveness of promoting health empowerment, online health information providers should establish a health literacy assessment mechanism to provide targeted information services to individuals with different health literacy. For individuals with a high level of health literacy, it is an effective strategy to cultivate users' health decision-making ability to promote health empowerment, and providers should highlight the scientific nature of the information. For those with a lower level of health literacy, making them more informed is an effective way to promote empowerment, and providers should highlight the professionalism and reliability of the sources of information.

Limitations of the Study

Although this paper draws some conclusions that cannot be ignored, there are still some shortcomings that should be addressed in the future. First, this study did not consider the impact of the type of online health information service model used by consumers to gather information, such as an online health consultation website or a medical information portal. The unique characteristics of different online health information services may impact health empowerment. Second, our study was based on a static model and cross-sectional data. The processes that affect the promotion of individual health empowerment are likely to be dynamic, so longitudinal research is necessary. Third, we did not involve the measurement of the respondent's disease and pathology, which may affect a person's use of online health information. Finally, medical consultation is a process of interaction between patients and doctors. This research only focuses on patient factors. Future research should consider the impact of doctor-related factors (such as empathy and patient-centered communication) on health empowerment.

Conclusion

In this paper, we explored the effect of the central route and peripheral route of online health information on users' health empowerment. We also considered the moderating role of health literacy in both routes. To test the hypothesis, PLS-SEM was used to analyze the data, and the empirical results supported most of the hypothesis. The findings further confirmed the important role of electronic information technology in promoting health empowerment. In the context of online health information, we must pay more attention to information quality and the interaction effect between individuals' health literacy and information processing cues. Research results provide practical guidance for health information providers to better serve and maximize individuals' benefits and empowerment. This study also pointed out the differences in promoting health empowerment besides health literacy. And more research in the future is needed to focus on individualized differences in the promotion of health empowerment.

Acknowledgments

This work was funded in part by the National Natural Science Foundation of China (grants 71771219 and 72071213). In addition, we appreciate the academic committee in the "Mobile Health" Ministry of Education-China Mobile Joint Laboratory for reviewing the project proposal and providing ethical approval.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement Scales.
[DOC File , 125 KB - jmir_v24i1e27178_app1.doc]

Multimedia Appendix 2

Common Method Bias Analysis.
[DOC File , 91 KB - jmir_v24i1e27178_app2.doc]

References

1. Castro EM, Van Regenmortel T, Vanhaecht K, Sermeus W, Van Hecke A. Patient empowerment, patient participation and patient-centeredness in hospital care: A concept analysis based on a literature review. Patient Educ Couns 2016 Dec;99(12):1923-1939. [doi: 10.1016/j.pec.2016.07.026] [Medline: 27450481]
2. Feste C, Anderson RM. Empowerment: from philosophy to practice. Patient Educ Couns 1995 Sep;26(1-3):139-144. [doi: 10.1016/0738-3991(95)00730-n] [Medline: 7494713]
3. Anderson RM, Funnell MM. Theory is the cart, vision is the horse: reflections on research in diabetes patient education. Diabetes Educ 1999;25(6 Suppl):43-51. [doi: 10.1177/014572179902500622] [Medline: 10711084]
4. Anderson RM, Funnell MM. Patient empowerment: reflections on the challenge of fostering the adoption of a new paradigm. Patient Educ Couns 2005 May;57(2):153-157. [doi: 10.1016/j.pec.2004.05.008] [Medline: 15911187]
5. The 46th China Statistical Report on Internet Development. CNNIC. 2020 Sep 29. URL: http://www.cnnic.net.cn/hlwfzyj/hlwxzbg/hlwtjbg [accessed 2021-12-16]
6. Almunawar M, Anshari M, Younis MZ. Incorporating customer empowerment in mobile health. Health Policy and Technology 2015 Dec;4(4):312-319. [doi: 10.1016/j.hlpt.2015.08.008]
7. Lemire M. What can be expected of information and communication technologies in terms of patient empowerment in health? J Health Organ Manag 2010;24(2):167-181. [doi: 10.1108/14777261011047336] [Medline: 20698192]
8. Vezyridis P, Timmons S. On the adoption of personal health records: some problematic issues for patient empowerment. Ethics Inf Technol 2015 Jun 26;17(2):113-124. [doi: 10.1007/s10676-015-9365-x]
9. Audrain-Pontevia A, Menvielle L. Do online health communities enhance patient-physician relationship? An assessment of the impact of social support and patient empowerment. Health Serv Manage Res 2018 Aug;31(3):154-162. [doi: 10.1177/0951484817748462] [Medline: 29280679]
10. Sharma S, Khadka A. Role of empowerment and sense of community on online social health support group. ITP 2019 Dec 02;32(6):1564-1590. [doi: 10.1108/itp-09-2018-0410]
11. van Uden-Kraan CF, Drossaert CHC, Taal E, Seydel ER, van de Laar MAFJ. Participation in online patient support groups endorses patients' empowerment. Patient Educ Couns 2009 Jan;74(1):61-69. [doi: 10.1016/j.pec.2008.07.044] [Medline: 18778909]
12. Petrič G, Atanasova S, Kamin T. Impact of Social Processes in Online Health Communities on Patient Empowerment in Relationship With the Physician: Emergence of Functional and Dysfunctional Empowerment. J Med Internet Res 2017 Mar 13;19(3):e74 [FREE Full text] [doi: 10.2196/jmir.7002] [Medline: 28288953]
13. van Berkel JJ, Lambooij MS, Hegger I. Empowerment of patients in online discussions about medicine use. BMC Med Inform Decis Mak 2015 Apr 08;15:24 [FREE Full text] [doi: 10.1186/s12911-015-0146-6] [Medline: 25888825]
14. Nelson EC, Verhagen T, Noordzij ML. Health empowerment through activity trackers: An empirical smart wristband study. Computers in Human Behavior 2016 Sep;62:364-374. [doi: 10.1016/j.chb.2016.03.065]
15. van den Berg SW, Gielissen MFM, Ottevanger PB, Prins JB. Rationale of the BREAst cancer e-healTH [BREATH] multicentre randomised controlled trial: an internet-based self-management intervention to foster adjustment after curative breast cancer by decreasing distress and increasing empowerment. BMC Cancer 2012 Sep 07;12:394 [FREE Full text] [doi: 10.1186/1471-2407-12-394] [Medline: 22958799]
16. Karni L, Dalal K, Memedi M, Kalra D, Klein GO. Information and Communications Technology-Based Interventions Targeting Patient Empowerment: Framework Development. J Med Internet Res 2020 Aug 26;22(8):e17459 [FREE Full text] [doi: 10.2196/17459] [Medline: 32845245]
17. McGloin H, O'Connell D, Glacken M, McSharry P, Healy D, Winters-O'Donnell L, et al. Patient Empowerment Using Electronic Telemonitoring With Telephone Support in the Transition to Insulin Therapy in Adults With Type 2 Diabetes: Observational, Pre-Post, Mixed Methods Study. J Med Internet Res 2020 May 14;22(5):e16161 [FREE Full text] [doi: 10.2196/16161] [Medline: 32406854]
18. Sanders K, Sánchez Valle M, Viñaras M, Llorente C. Do we trust and are we empowered by “Dr. Google”? Older Spaniards’ uses and views of digital healthcare communication. Public Relations Review 2015 Dec;41(5):794-800. [doi: 10.1016/j.pubrev.2015.06.015]
19. Johnston AC, Worrell JL, Di Gangi PM, Wasko M. Online health communities: An assessment of the influence of participation on patient empowerment outcomes. Info Technology & People 2013 May 31;26(2):213-235. [doi: 10.1108/ITP-02-2013-0040]
20. Pecukonis E. Health literacy: A prescription to end confusion. Social Work in Health Care 2008 Sep 21;46(4):101-106. [doi: 10.1300/J010v46n04_06]
21. Meppelink CS, Smit EG, Diviani N, Van Weert JCM. Health Literacy and Online Health Information Processing: Unraveling the Underlying Mechanisms. J Health Commun 2016;21(sup2):109-120. [doi: 10.1080/10810730.2016.1193920] [Medline: 27668318]
22. Gibson CH. A concept analysis of empowerment. J Adv Nurs 1991 Mar;16(3):354-361. [doi: 10.1111/j.1365-2648.1991.tb01660.x] [Medline: 2037742]

23. Nyatanga L, Dann KL. Empowerment in nursing: the role of philosophical and psychological factors. Nursing Philosophy 2002 Oct 21;3(3):234-239. [doi: 10.1046/j.1466-769X.2002.00107.x]
24. Labonte R. Community and professional empowerment. Can Nurse 1989 Mar;85(3):22-28. [Medline: 2924275]
25. Anderson RM, Funnell MM. Patient empowerment: myths and misconceptions. Patient Educ Couns 2010 Jun;79(3):277-282 [FREE Full text] [doi: 10.1016/j.pec.2009.07.025] [Medline: 19682830]
26. Seçkin G. Cyber patients surfing the medical web: Computer-mediated medical knowledge and perceived benefits. Computers in Human Behavior 2010 Nov;26(6):1694-1700. [doi: 10.1016/j.chb.2010.06.018]
27. Lee Y, Shin S, Wang R, Lin K, Lee Y, Wang Y. Pathways of empowerment perceptions, health literacy, self-efficacy, and self-care behaviors to glycemic control in patients with type 2 diabetes mellitus. Patient Educ Couns 2016 Feb;99(2):287-294. [doi: 10.1016/j.pec.2015.08.021] [Medline: 26341940]
28. Moattari M, Ebrahimi M, Sharifi N, Rouzbeh J. The effect of empowerment on the self-efficacy, quality of life and clinical and laboratory indicators of patients treated with hemodialysis: a randomized controlled trial. Health Qual Life Outcomes 2012 Sep 20;10:115 [FREE Full text] [doi: 10.1186/1477-7525-10-115] [Medline: 22992449]
29. Fumagalli LP, Radaelli G, Lettieri E, Bertele' P, Masella C. Patient Empowerment and its neighbours: clarifying the boundaries and their mutual relationships. Health Policy 2015 Mar;119(3):384-394. [doi: 10.1016/j.healthpol.2014.10.017] [Medline: 25467286]
30. Funnell MM, Anderson RM. Empowerment and Self-Management of Diabetes. Clinical Diabetes 2004 Jul 01;22(3):123-127. [doi: 10.2337/diaclin.22.3.123]
31. Newton P, Sasha S, Koula A. Marrying contradictions: healthcare professionals perceptions of empowerment in the care of people with Type 2 Diabetes. Patient Educ Couns 2011 Dec;85(3):e326-e329. [doi: 10.1016/j.pec.2011.03.015] [Medline: 21530141]
32. Funnell MM, Anderson RM, Arnold MS, Barr PA, Donnelly M, Johnson PD, et al. Empowerment: an idea whose time has come in diabetes education. Diabetes Educ 1991;17(1):37-41. [doi: 10.1177/014572179101700108] [Medline: 1986902]
33. Ouschan R, Sweeney J, Johnson L. Customer empowerment and relationship outcomes in healthcare consultations. European Journal of Marketing 2006 Sep;40(9/10):1068-1086. [doi: 10.1108/03090560610681014]
34. Prigge J, Dietz B, Homburg C, Hoyer WD, Burton JL. Patient empowerment: A cross-disease exploration of antecedents and consequences. International Journal of Research in Marketing 2015 Dec;32(4):375-386. [doi: 10.1016/j.ijresmar.2015.05.009]
35. Londoño AMM, Schulz PJ. Influences of health literacy, judgment skills, and empowerment on asthma self-management practices. Patient Educ Couns 2015 Jul;98(7):908-917. [doi: 10.1016/j.pec.2015.03.003] [Medline: 25907737]
36. Bann CM, Sirois FM, Walsh EG. Provider support in complementary and alternative medicine: exploring the role of patient empowerment. J Altern Complement Med 2010 Jul;16(7):745-752. [doi: 10.1089/acm.2009.0381] [Medline: 20575703]
37. Kim SC, Kim S, Boren D. The quality of therapeutic alliance between patient and provider predicts general satisfaction. Mil Med 2008 Jan;173(1):85-90. [doi: 10.7205/milmed.173.1.85] [Medline: 18251337]
38. Petty RE, Cacioppo JT, Goldman R. Personal involvement as a determinant of argument-based persuasion. Journal of Personality and Social Psychology 1981;41(5):847-855. [doi: 10.1037//0022-3514.41.5.847]
39. Watts S, Zhang W. Capitalizing on Content: Information Adoption in Two Online communities. JAIS 2008 Feb;9(2):73-94. [doi: 10.17705/1jais.00149]
40. Sussman SW, Siegal WS. Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption. Information Systems Research 2003 Mar;14(1):47-65. [doi: 10.1287/isre.14.1.47.14767]
41. Cao X, Liu Y, Zhu Z, Hu J, Chen X. Online selection of a physician by patients: Empirical study from elaboration likelihood perspective. Computers in Human Behavior 2017 Aug;73:403-412. [doi: 10.1016/j.chb.2017.03.060]
42. Angst, Agarwal. Adoption of Electronic Health Records in the Presence of Privacy Concerns: The Elaboration Likelihood Model and Individual Persuasion. MIS Quarterly 2009 Jun;33(2):339-370. [doi: 10.2307/20650295]
43. Bhattacherjee, Sanford. Influence Processes for Information Technology Acceptance: An Elaboration Likelihood Model. MIS Quarterly 2006;30(4):805-825. [doi: 10.2307/25148755]
44. Yi MY, Yoon JJ, Davis JM, Lee T. Untangling the antecedents of initial trust in Web-based health information: The roles of argument quality, source expertise, and user perceptions of information quality and risk. Decision Support Systems 2013 Apr;55(1):284-295. [doi: 10.1016/j.dss.2013.01.029]
45. Marquart F, Naderer B. Communication and Persuasion: Central and Peripheral Routes to Attitude Change. In: Potthoff M, editor. Schlüsselwerke der Medienwirkungsforschung. Wiesbaden: Springer VS; 2016:231-242.
46. Wang HC, Doong HS, Shih HC, Pallister J, Foxall G. An Investigation into the determinants of repurchase loyalty in the E-Marketplace. 2008 Presented at: Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS 2008); 7-10 Jan. 2008; Waikoloa, HI, USA p. 272-272. [doi: 10.1109/HICSS.2008.62]
47. Koo C, Wati Y. E-Healthcare Service: An Investigation of the Antecedents, Moderating Roles, and Consequences. 2011 Presented at: 44th Hawaii International Conference on System Sciences; 4-7 Jan. 2011; Kauai, HI, USA p. 1-10. [doi: 10.1109/HICSS.2011.176]
48. Boller GW, Swasy JL, Munch JM. Conceptualizing argument quality via argument structure. Advances in Consumer Research 1990;17(1):321-328.

49. Pornpitakpan C. The Persuasiveness of Source Credibility: A Critical Review of Five Decades' Evidence. J Appl Social Pyschol 2004 Feb;34(2):243-281. [doi: 10.1111/j.1559-1816.2004.tb02547.x]
50. Harrison JP, Lee A. The role of e-Health in the changng health care environment. Nurs Econ 2006;24(6):283-8, 279; quiz 289. [Medline: 17266004]
51. Zaphiris P, Ang CS, editors. Social Computing and Virtual Communities. Boca Raton, FL: Chapman & Hall/CRC; 2009.
52. Watanabe N, Kaneko A, Yamar S, Taleo G, Tanihata T, Lum JK, et al. A prescription for sustaining community engagement in malaria elimination on Aneityum Island, Vanuatu: an application of Health Empowerment Theory. Malar J 2015 Jul 31;14:291-301 [FREE Full text] [doi: 10.1186/s12936-015-0779-z] [Medline: 26228787]
53. Fleisher L, Bass S, Ruzek SB, McKeown-Conn N. Relationships among Internet health information use, patient behavior and self efficacy in newly diagnosed cancer patients who contact the National Cancer Institute's NCI Atlantic Region Cancer Information Service (CIS). Proc AMIA Symp 2002:260-264 [FREE Full text] [Medline: 12463827]
54. Ouschan R, Sweeney JC, Johnson LW. Dimensions of patient empowerment: implications for professional services marketing. Health Mark Q 2000;18(1-2):99-114. [doi: 10.1300/J026v18n01_08] [Medline: 11184439]
55. Hain DJ, Sandy D. Partners in care: patient empowerment through shared decision-making. Nephrol Nurs J 2013;40(2):153-157. [Medline: 23767339]
56. Williams T. Patient empowerment and ethical decision making: the patient/partner and the right to act. Dimens Crit Care Nurs 2002;21(3):100-104. [doi: 10.1097/00003465-200205000-00006] [Medline: 12042695]
57. Clarke E, Puschner B, Jordan H, Williams P, Konrad J, Kawohl W, et al. Empowerment and satisfaction in a multinational study of routine clinical practice. Acta Psychiatr Scand 2015 May;131(5):369-378. [doi: 10.1111/acps.12365] [Medline: 25471821]
58. Areni CS, Lutz RJ. The role of argument quality in the elaboration likelihood model. Advances in Consumer Research 1988;15(1):197-203.
59. Chaiken S, Eagly AH. Communication modality as a determinant of message persuasiveness and message comprehensibility. Journal of Personality and Social Psychology 1976 Oct;34(4):605-614. [doi: 10.1037/0022-3514.34.4.605]
60. Lin T, Hwang L, Lai Y. Effects of argument quality, source credibility and self-reported diabetes knowledge on message attitudes: an experiment using diabetes related messages. Health Info Libr J 2017 Sep;34(3):225-235 [FREE Full text] [doi: 10.1111/hir.12181] [Medline: 28513005]
61. Raghunathan S. Impact of information quality and decision-maker quality on decision quality: a theoretical model and simulation analysis. Decision Support Systems 1999 Oct;26(4):275-286. [doi: 10.1016/S0167-9236(99)00060-3]
62. Chen C, C.S. Ku E. Bridging indistinct relationships and online loyalty: evidence from online interest-based communities. Online Information Review 2013 Sep 23;37(5):731-751. [doi: 10.1108/oir-01-2011-0220]
63. Tormala ZL, Briñol P, Petty RE. When credibility attacks: The reverse impact of source credibility on persuasion. Journal of Experimental Social Psychology 2006 Sep;42(5):684-691. [doi: 10.1016/j.jesp.2005.10.005]
64. Zhu DH, Chang YP, Luo JJ. Understanding the influence of C2C communication on purchase decision in online communities from a perspective of information adoption model. Telematics and Informatics 2016 Feb;33(1):8-16. [doi: 10.1016/j.tele.2015.06.001]
65. Syn SY, Kim SU. The impact of source credibility on young adults' Health information activities on facebook: Preliminary findings. Proc. Am. Soc. Info. Sci. Tech 2014 May 08;50(1):1-4. [doi: 10.1002/meet.14505001122]
66. Xiao N, Sharman R, Rao H, Upadhyaya S. Factors influencing online health information search: An empirical analysis of a national cancer-related survey. Decision Support Systems 2014 Jan;57:417-427. [doi: 10.1016/j.dss.2012.10.047]
67. Ghaddar SF, Valerio MA, Garcia CM, Hansen L. Adolescent health literacy: the importance of credible sources for online health information. J Sch Health 2012 Jan;82(1):28-36. [doi: 10.1111/j.1746-1561.2011.00664.x] [Medline: 22142172]
68. Bates BR, Romina S, Ahmed R, Hopson D. The effect of source credibility on consumers' perceptions of the quality of health information on the Internet. Med Inform Internet Med 2006 Mar;31(1):45-52. [doi: 10.1080/14639230600552601] [Medline: 16754366]
69. Lin CA, Atkin DJ, Cappotto C, Davis C, Dean J, Eisenbaum J, et al. Ethnicity, digital divides and uses of the Internet for health information. Computers in Human Behavior 2015 Oct;51:216-223. [doi: 10.1016/j.chb.2015.04.054]
70. Pant S, Deshmukh A, Murugiah K, Kumar G, Sachdeva R, Mehta JL. Assessing the credibility of the "YouTube approach" to health information on acute myocardial infarction. Clin Cardiol 2012 May;35(5):281-285 [FREE Full text] [doi: 10.1002/clc.21981] [Medline: 22487995]
71. Wei M. The associations between health literacy, reasons for seeking health information, and information sources utilized by Taiwanese adults. Health Education Journal 2013 Mar 26;73(4):423-434. [doi: 10.1177/0017896912471523]
72. Tennant B, Stellefson M, Dodd V, Chaney B, Chaney D, Paige S, et al. eHealth literacy and Web 2.0 health information seeking behaviors among baby boomers and older adults. J Med Internet Res 2015 Mar 17;17(3):e70 [FREE Full text] [doi: 10.2196/jmir.3992] [Medline: 25783036]
73. Slim K, Bazin J. From informed consent to shared decision-making in surgery. J Visc Surg 2019 Jun;156(3):181-184. [doi: 10.1016/j.jviscsurg.2019.04.014] [Medline: 31101549]
74. Chang H, Li F, Lin C. Factors Influencing Implementation Of Shared Medical Decision Making In Patients With Cancer. PPA 2019 Nov;13:1995-2005. [doi: 10.2147/ppa.s217561]

75. Seo J, Goodman MS, Politi M, Blanchard M, Kaphingst KA. Effect of Health Literacy on Decision-Making Preferences among Medically Underserved Patients. Med Decis Making 2016 May;36(4):550-556 [FREE Full text] [doi: 10.1177/0272989X16632197] [Medline: 26902737]
76. Hur K, Kim TT, Karatepe OM, Lee G. An exploration of the factors influencing social media continuance usage and information sharing intentions among Korean travellers. Tourism Management 2017 Dec;63:170-178. [doi: 10.1016/j.tourman.2017.06.013]
77. Sudbury-Riley L, FitzPatrick M, Schulz PJ. Exploring the Measurement Properties of the eHealth Literacy Scale (eHEALS) Among Baby Boomers: A Multinational Test of Measurement Invariance. J Med Internet Res 2017 Feb 27;19(2):e53 [FREE Full text] [doi: 10.2196/jmir.5998] [Medline: 28242590]
78. Foley H, Steel A. Patient perceptions of patient-centred care, empathy and empowerment in complementary medicine clinical practice: A cross-sectional study. Advances in Integrative Medicine 2017 Apr;4(1):22-30. [doi: 10.1016/j.aimed.2017.02.006]
79. Krause N, Riemann-Lorenz K, Steffen T, Rahn AC, Pöttgen J, Stellmann J, et al. Study protocol for a randomised controlled trial of a web-based behavioural lifestyle programme for emPOWERment in early Multiple Sclerosis (POWER@MS1). BMJ Open 2021 Feb 16;11(2):e041720 [FREE Full text] [doi: 10.1136/bmjopen-2020-041720] [Medline: 33593774]
80. Rahn AC, Wenzel L, Icks A, Stahmann A, Scheiderbauer J, Grentzenberg K, et al. Development and evaluation of an interactive web-based decision-making programme on relapse management for people with multiple sclerosis (POWER@MS2)-study protocol for a randomised controlled trial. Trials 2021 Feb 14;22(1):139 [FREE Full text] [doi: 10.1186/s13063-021-05059-1] [Medline: 33583424]
81. Changsha Ranxing Information Technology Co., Ltd. URL: https://www.wjx.cn/ [accessed 2021-12-15]
82. Hair Jr. JF, Matthews LM, Matthews RL, Sarstedt M. PLS-SEM or CB-SEM: updated guidelines on which method to use. IJMDA 2017;1(2):107-123. [doi: 10.1504/ijmda.2017.10008574]
83. Ringle, Sarstedt, Straub. Editor's Comments: A Critical Look at the Use of PLS-SEM in "MIS Quarterly". MIS Quarterly 2012;36(1):iii-xiv. [doi: 10.2307/41410402]
84. Hair JF, Sarstedt M, Ringle CM, Mena JA. An assessment of the use of partial least squares structural equation modeling in marketing research. J Acad Mark Sci 2011 Jun 7;40(3):414-433. [doi: 10.1007/s11747-011-0261-6]
85. Fornell C, Larcker D. Evaluating Structural Equation Models with Unobservable Variables and Measurement Error. Journal of Marketing Research 1981 Feb 1;18(1):39-50. [doi: 10.1177/002224378101800104]
86. Bagozzi RP, Yi Y. On the evaluation of structural equation models. JAMS 1988 Mar;16(1):74-94. [doi: 10.1007/bf02723327]
87. Podsakoff PM, Organ DW. Self-Reports in Organizational Research: Problems and Prospects. Journal of Management 1986 Dec 1;12(4):531-544. [doi: 10.1177/014920638601200408]
88. Liang, Saraf, Hu, Xue. Assimilation of Enterprise Systems: The Effect of Institutional Pressures and the Mediating Role of Top Management. MIS Quarterly 2007;31(1):59-87. [doi: 10.2307/25148781]
89. Keil M, Tan BCY, Wei K, Saarinen T, Tuunainen V, Wassenaar A. A Cross-Cultural Study on Escalation of Commitment Behavior in Software Projects. MIS Quarterly 2000 Jun;24(2):299-325. [doi: 10.2307/3250940]
90. Zhang X, Ma L, Xu B, Xu F. How social media usage affects employees' job satisfaction and turnover intention: An empirical study in China. Information & Management 2019 Sep;56(6):103136. [doi: 10.1016/j.im.2018.12.004]
91. Hwang Y. The moderating effects of gender on e-commerce systems adoption factors: An empirical investigation. Computers in Human Behavior 2010 Nov;26(6):1753-1760. [doi: 10.1016/j.chb.2010.07.002]
92. Zhao X, Lynch JG, Chen Q. Reconsidering Baron and Kenny: Myths and Truths about Mediation Analysis. J Consum Res 2010 Aug 01;37(2):197-206. [doi: 10.1086/651257]
93. MacKinnon DP, Lockwood CM, Hoffman JM, West SG, Sheets V. A comparison of methods to test mediation and other intervening variable effects. Psychol Methods 2002 Mar;7(1):83-104 [FREE Full text] [doi: 10.1037/1082-989x.7.1.83] [Medline: 11928892]
94. Shin KS, Lee E. Relationships of health literacy to self-care behaviors in people with diabetes aged 60 and above: Empowerment as a mediator. J Adv Nurs 2018 Oct 05;74(10):2363-2372. [doi: 10.1111/jan.13738] [Medline: 29893030]
95. Crondahl K, Eklund Karlsson L. The Nexus Between Health Literacy and Empowerment. SAGE Open 2016 May 02:1-7. [doi: 10.1177/2158244016646410]
96. Finbråten HS, Guttersrud, Nordström G, Pettersen KS, Trollvik A, Wilde-Larsson B. Explaining variance in health literacy among people with type 2 diabetes: the association between health literacy and health behaviour and empowerment. BMC Public Health 2020 Feb 03;20(1):161 [FREE Full text] [doi: 10.1186/s12889-020-8274-z] [Medline: 32013897]

Abbreviations

AVE: average variance extracted
CMV: common method variance
ELM: elaboration likelihood model
ICT: information and communication technology
PLS-SEM: partial least squares structural equation modeling

Edited by R Kukafka; submitted 14.01.21; peer-reviewed by L Menvielle, C Meppelink, J Rowley; comments to author 02.03.21; revised version received 31.03.21; accepted 04.11.21; published 11.01.22.

Please cite as:
Jiang F, Liu Y, Hu J, Chen X
Understanding Health Empowerment From the Perspective of Information Processing: Questionnaire Study
J Med Internet Res 2022;24(1):e27178
URL: https://www.jmir.org/2022/1/e27178
doi:10.2196/27178
PMID:35014957

Original Paper

A Remote Monitoring System to Optimize the Home Management of Oral Anticancer Therapies (ONCO-TreC): Prospective Training–Validation Trial

Alessandro Passardi[1], MD; Flavia Foca[2], MStat; Orazio Caffo[3], MD; Carlo Alberto Tondini[4], MD; Alberto Zambelli[4], MD; Roberto Vespignani[5], BEng; Giulia Bartolini[1], MD; Francesco Giulio Sullo[1], MD; Daniele Andreis[2], BSc; Marco Dianti[6], MSc; Claudio Eccher[6], PhD; Enrico Maria Piras[6], PhD; Stefano Forti[6], MSc

[1]Department of Medical Oncology, IRCCS Istituto Romagnolo per lo Studio dei Tumori "Dino Amadori", Meldola, Italy

[2]Unit of Biostatistics and Clinical Trials, IRCCS Istituto Romagnolo per lo Studio dei Tumori "Dino Amadori", Meldola, Italy

[3]Department of Medical Oncology, Azienda Provinciale per i Servizi Sanitari, Trento, Italy

[4]Department of Medical Oncology, Azienda Socio-Sanitaria Territoriale "Papa Giovanni XXIII", Bergamo, Italy

[5]IT Service, IRCCS Istituto Romagnolo per lo Studio dei Tumori "Dino Amadori", Meldola, Italy

[6]Center for Information and Communication Technology, eHealth Unit, Fondazione "Bruno Kessler", Trento, Italy

Corresponding Author:
Roberto Vespignani, BEng
IT Service
IRCCS Istituto Romagnolo per lo Studio dei Tumori "Dino Amadori"
Via P. Maroncelli 40
Meldola, 47014
Italy
Phone: 39 0543 739992
Email: roberto.vespignani@irst.emr.it

Abstract

Background: A platform designed to support the home management of oral anticancer treatments and provide a secure web-based patient–health care professional communication modality, ONCO-TreC, was tested in 3 cancer centers in Italy.

Objective: The overall aims of the trial are to customize the platform; assess the system's ability to facilitate the shared management of oral anticancer therapies by patients and health professionals; and evaluate system usability and acceptability by patients, caregivers, and health care professionals.

Methods: Patients aged ≥18 years who were candidates for oral anticancer treatment as monotherapy with an Eastern Cooperative Oncology Group performance status score of 0 to 1 and a sufficient level of familiarity with mobile devices were eligible. ONCO-TreC consisted of a mobile app for patients and a web-based dashboard for health care professionals. Adherence to treatment (pill count) and toxicities reported by patients through the app were compared with those reported by physicians in medical records. Usability and acceptability were evaluated using questionnaires.

Results: A total of 40 patients were enrolled, 38 (95%) of whom were evaluable for adherence to treatment. The ability of the system to measure adherence to treatment was high, with a concordance of 97.3% (95% CI 86.1%-99.9%) between the investigator and system pill count. Only 60% (3/5) of grade 3, 54% (13/24) of grade 2, and 19% (7/36) of grade 1 adverse events reported by physicians in the case report forms were also reported in the app directly by patients. In total, 94% (33/35) of patients had ≥1 app launch each week, and the median number of daily accesses per patient was 2. Approximately 71% (27/38) and 68% (26/38) of patients used the app for messages and vital sign entering, respectively, at least once during the study period.

Conclusions: ONCO-TreC is an important tool for measuring and monitoring adherence to oral anticancer drugs. System usability and acceptability were very high, whereas its reliability in registering toxicity could be improved.

Trial Registration: ClinicalTrials.gov NCT02921724; https://www.clinicaltrials.gov/ct2/show/NCT02921724

(J Med Internet Res 2022;24(1):e27349) doi:10.2196/27349

KEYWORDS

adherence; oral anticancer drug; mHealth; ONCO-TreC; electronic diary

Introduction

During the last few decades, oral anticancer drugs, either alone or in combination with intravenous treatments, have occupied an increasingly important space in oncohematology. In addition to traditional hormonal and cytotoxic drugs, new formulations targeting specific genetic mutations (usually referred to as tyrosine kinase inhibitors [TKIs]) have been widely developed [1,2]. The advantages of oral therapies include improved outcomes in several tumor types and a reduction in the workload needed for nurses in terms of administration and intravenous infusion. Moreover, patients generally prefer this type of administration as it enables them to maintain their normal lifestyle [3]. However, alongside these indisputable benefits, some critical issues regarding the use of oral treatments have emerged, especially in relation to treatment adherence and patient safety [4]. For example, toxicities may not be reported, major drug interactions may be overlooked, and self-administration may expose patients to the risk of over- or undertreatment.

Several studies addressing adherence to oral treatments have reported variable rates, some as low as 6% [5-9]. Moreover, adherence levels appear to influence specific clinical and health care outcomes, such as cancer progression, inpatient days, health care resource use and costs, and even survival [10-15]. Numerous variables have been correlated with nonadherence in relation to patients (eg, age and beliefs about medication), treatments (eg, toxicity and complex schedules), health care professionals (eg, empathy and communication skills), and the health care system (eg, communication problems with cancer centers) [16-18]. Consequently, methods that can be used to increase adherence levels include patient education and improved communication between health care professionals and patients [19,20].

Patient-centered approaches and mobile health care solutions such as web-based and mobile apps have proven to be useful tools for optimizing the home management of cancer patients. In this context, by involving patients and health care professionals in participatory design techniques (ie, focus group sessions and joint reviews), we were able to customize and adopt an existing health care system (already used for the remote monitoring of patients with asthma, type 1 diabetes, and hypertension) that met the needs of home management and remote monitoring of oral anticancer therapy [21]. We have conducted a prospective training–validation, interventional, nonpharmacological, multicenter study on a new platform called ONCO-TreC for oral anticancer therapy. The overall aims of the trial are to customize the platform; assess the ability of the system to support patients and health care professionals in the shared management of oral anticancer therapies (thus improving adherence and the home management of side effects); and evaluate the usability and acceptability of the system by patients, caregivers, and health care professionals.

Methods

ONCO-TreC

ONCO-TreC consists of a mobile app for patients and a web-based dashboard for health care professionals. Clinicians enter the details of medication schemes through the dashboard, set reminders, monitor for adherence to treatment and adverse events (AEs), and communicate with patients via a secure messaging system. The app provides patients with a visual reminder of cancer therapy, a reminder of concomitant drugs to be taken, an easy-to-use AE reporting system, a diary of vital signs, and a messaging system. A detailed description of ONCO-TreC and the trial design has been previously reported [22]. All the cancer centers involved in the study used electronic health records, and although the ONCO-TreC system was designed to be fully integrated into the health records, it was still not connected at this experimental stage. Thus, the investigators consulted the medical records and dashboard separately.

Patients

A total of 80 patients were considered for the ONCO-TreC trial, comprising 25% (20/80) of patients in the training step and 75% (60/80) patients in the validation step. Eligible patients were required to meet the following criteria: (1) adults aged ≥18 and <75 years of either sex, (2) Eastern Cooperative Oncology Group performance status score of ≤1, (3) candidates for oral anticancer treatment with capecitabine or sunitinib as monotherapy (adjuvant and advanced settings allowed), (4) ability to manage the mobile app after a basic training course held at baseline, (5) clear understanding of the Italian language, and (6) written informed consent. Patients who were also receiving intravenous anticancer treatment or experimental drugs were excluded.

At the time of treatment prescription, the health care professionals provided information on treatment-related side effects and the use and functions of the app. The patients were advised to manually insert data into the system at least once a day. They were seen at study centers every 6 weeks, when clinicians compared adherence and toxicity data entered into the system with those directly reported during the hospital visit and with drug accountability. If the patients were having technical difficulties with the mobile app, the training was repeated. The patients remained under observation until treatment interruption (because of disease progression, unacceptable toxicity, death, or discontinuation) or for ≤6 months.

At the end of the training step (first 20 patients), the results were shared among the centers, the system was fine-tuned, and the study protocol was amended. The sample size of the validation step was reduced from 60 to 20 patients, which is considered a sufficient number to validate the system in a clinical setting. The inclusion criteria were expanded, allowing for the inclusion of patients aged >75 years and those being treated with other

TKIs (regorafenib, sorafenib, pazopanib, everolimus, erlotinib, gefitinib, afatinib, axitinib, or crizotinib). The decision to reduce the sample size of the validation step from 60 to 20 patients was made because the analysis of the training step had already highlighted excellent usability and high patient satisfaction. Conversely, to have more efficacy data, it was deemed necessary to design a larger randomized study comparing the effectiveness of the system with that of standard clinical practice.

Outcome Measures

To assess the system's ability to monitor adherence, the number of pills counted by the system (entered by the patients at home) was compared with that of residual pills returned by the patients during the hospital visit and counted by the physicians. Patients who took ≥90% of the total drug dose as per the study protocol were defined as adherent. The proportions of adherent patients according to the app and to the pill count were compared. AEs were reported and graded on a daily basis by patients through the app, whereas during study visits, the oncologist recorded the highest grade of each AE per cycle in medical records. The reliability of the system for registering toxicity was evaluated by comparing the type and grade of toxicity indicated by the system with those registered during the clinical visits. The quality of the system was considered adequate if all grade 3 toxicities and ≥80% of grade 2 toxicities reported by the patient at the time of the visit were recorded in the app. In total, 2 validated questionnaires, 1 on perceived levels of quality of life (Functional Assessment of Cancer Therapy-General [FACT-G]) and the other on anxiety (Hospital Anxiety and Depression Scale [HADS]), were administered at baseline and at the end of the study to assess health-related quality of life and anxiety and depression levels in the enrolled patients [23,24]. Given that a review of the literature did not identify any existing validated questionnaires for assessing system usability and acceptability by patients, 2 ad hoc questionnaires were developed by the team of investigators to analyze patient expectations with regard to the system (administered at baseline) and to evaluate system acceptability and communication between patients and cancer centers (administered at week 6 and at the end of treatment). The first questionnaire aimed to establish a baseline with regard to patient communication with the oncology department in terms of the means used (eg, frequency of emails or phone calls) and the adequacy of responses to questions asked. The second questionnaire, distributed at the end of the study, aimed to assess the changes brought about by the use of ONCO-TreC. The overall usability of the system was assessed using the System Usability Scale (SUS) questionnaire, a validated 10-item Likert scale that provides a rapid, reliable subjective evaluation [25]. The SUS was administered at the end of the study to ensure that the patients had had sufficient time to familiarize themselves with the digital device. All the questionnaires were paper-based and self-completed by patients.

Statistical Analysis

A formal sample size calculation for the prospective study was not performed owing to the lack of preliminary data and the exploratory intent of the study. Frequencies were calculated for categorical variables. For continuous variables, median (minimum to maximum or IQR) or mean and SD were shown. The Wilcoxon and Fisher tests were used to evaluate the difference between the ratios of days of use to total days in different groups of patients. The following tests were used for the comparison between the baseline and end-of-study questionnaires: Wilcoxon matched-pairs signed-ranks for the FACT-G and Stuart-Maxwell for the HADS.

Results

System Fine-tuning

The study was activated on May 29, 2015. The TreC system, originally designed for remote monitoring of patients with asthma, type 1 diabetes, and hypertension, was customized to meet the home management and remote monitoring needs of patients with cancer treated with the cytotoxic drug capecitabine or the TKI sunitinib. From June 2015 to December 2015, the system, in particular the mobile app for patients, was fine-tuned through participatory research to comply with the clinical practice regulations of the participating centers. Face-to-face participatory design sessions were conducted with study researchers and patient representatives, the results of which were used to modify the system. Of note, a shared revision of the Common Terminology Criteria for Adverse Events version 4.03, based on health literacy and patient-reported outcome principles, was included in the system to support patient self-reporting of AEs (Multimedia Appendix 1). Some customized suggestions for the management of side effects, differentiated according to the degree of toxicity, were also implemented.

Patients

From January 2016 to July 2018, a total of 40 patients (20/40, 50% in the training step and 20/40, 50% in the validation step) were enrolled in the ONCO-TreC study from 3 cancer centers in different Italian regions. Patient characteristics are reported in Table 1. The median age was 66 years (range 42-82 years), and 8 patients were aged >75 years. Most were women (24/40, 60%) and had an Eastern Cooperative Oncology Group performance status score of 0 (31/40, 78%) and ≥1 comorbidity (32/40, 80%). As expected, cardiovascular and metabolic or endocrine comorbidities were the most frequent (22/40, 55% and 16/40, 40%, respectively). Oral anticancer drugs comprised mainly capecitabine (23/40, 58%), regorafenib (7/40, 18%), and sunitinib (6/40, 15%). Approximately 68% (27/40) of patients were being treated for advanced disease, and 44% (16/40) had been heavily pretreated. The patients used the app for a mean of 4.4 (SD 8.0) months.

Table 1. Patient characteristics (N=40).

Variable	Values
Age at study registration (years), median (range)	66 (42-82)
Gender, n (%)	
Male	16 (40)
Female	24 (60)
ECOG[a] (performance status), n (%)	
0	31 (78)
1	9 (23)
Comorbidity, n (%)	
Yes	32 (80)
No	8 (20)
Type of comorbidity, n (%)	
Cardiovascular	22 (55)
Pulmonary	4 (10)
Gastrointestinal and hepatobiliary	7 (18)
Metabolic and endocrine	16 (40)
Musculoskeletal	2 (5)
Renal or urinary tract	3 (8)
Allergy	4 (10)
Neurological and psychiatric	5 (13)
Other comorbidities	4 (10)
Site of disease, n (%)	
Colorectum	15 (38)
Pancreas	5 (13)
Lung	3 (8)
Breast	4 (10)
Biliary tract	2 (5)
Kidney	5 (13)
Liver	1 (3)
Unknown	5 (13)
Oral anticancer drug, n (%)	
Capecitabine	23 (58)
Regorafenib	7 (18)
Sunitinib	6 (15)
Afatinib	2 (5)
Sorafenib	1 (3)
Gefitinib	1 (3)
Setting, n (%)	
Adjuvant	10 (25)
Advanced	30 (75)
Previous therapy, n (%)	
None	14 (35)
1	6 (15)

Variable	Values
≥2	16 (40)
Unknown	4 (10)

[a]ECOG: Eastern Cooperative Oncology Group.

Treatment Adherence

Adherence to treatment, calculated according to pill count and registered in a case report form (CRF), was available for 38 patients. Of those 38 patients, 32 (84%) patients were defined as adherent to treatment, 1 (3%) patient had a residual drug that was justified (treatment discontinuation advised by the oncologist because of toxicity), and 5 (13%) patients were defined as nonadherent to treatment. Adherence to treatment according to ONCO-TreC was as follows: 87% (33/38) of patients were adherent, 3% (1/38) were justified nonadherent, and 11% (4/38) were nonadherent. A patient in the nonadherent group reported regular drug intake in the app even during the days of discontinuation. A concordance of 97.3% (95% CI 86.1%-99.9%) was observed between the CRF and app-reported adherence.

Toxicity Monitoring

Of the 40 enrolled patients, 35 (88%) were evaluable for toxicity. A total of 718 AEs were registered on ONCO-TreC by 18 patients, and 98 AEs were registered for 29 patients by physicians. In total, 1 patient reported AEs in the system but not during the hospital visits. Conversely, of the 29 patients reporting AEs during the study visits, 12 (41%) did not enter any AE information into the system. Table 2 summarizes the AEs reported by patients during study visits and registered in CRFs and those entered by the patients themselves into the app based on grade of severity. Of the 5 grade 3 AEs, 24 grade 2 AEs, and 36 grade 1 AEs reported by physicians in CRFs, 3 (60%), 13 (54%), and 7 (19%) cases, respectively, were also reported in the system by patients.

Table 2. Summary of adverse events reported in case report forms (CRF) or the app by grade of severity (N=35).

Adverse event	Registered events in the app, n (%)			Registered events in CRFs, n (%)		
	Grade 1 (n=45)	Grade 2 (n=26)	Grade 3 (n=4)	Grade 1 (n=36)	Grade 2 (n=24)	Grade 3 (n=5)
Asthenia or fatigue	5 (14)	5 (14)	1 (3)	3 (8)	7 (19)	1 (3)
Nausea or vomiting	14 (40)	5 (14)	0 (0)	4 (11)	8 (23)	1 (3)
Rash	1 (3)	0 (0)	0 (0)	1 (3)	0 (0)	0 (0)
Paronychia	0 (0)	0 (0)	0 (0)	1 (3)	0 (0)	0 (0)
Anorexia	1 (3)	1 (3)	1 (3)	4 (11)	1 (3)	0 (0)
Mucositis	3 (8)	3 (8)	1 (3)	1 (3)	3 (8)	1 (3)
Skin toxicity	3 (8)	2 (6)	0 (0)	3 (8)	1 (3)	0 (0)
Nervous system disorders	1 (3)	0 (0)	0 (0)	3 (8)	0 (0)	0 (0)
Alopecia	2 (6)	2 (6)	0 (0)	1 (3)	0 (0)	0 (0)
Hand and foot syndrome	3 (8)	3 (8)	1 (3)	4 (11)	2 (6)	2 (6)
Conjunctivitis	2 (6)	0 (0)	0 (0)	1 (3)	0 (0)	0 (0)
Diarrhea	8 (23)	4 (11)	0 (0)	5 (14)	2 (3)	0 (0)
Fever	0 (0)	0 (0)	0 (0)	3 (8)	0 (0)	0 (0)
Pruritus	2 (6)	1 (3)	0 (0)	2 (6)	0 (0)	0 (0)

System Use

A total of 5186 accesses to the app were made by 35 patients. Of these35 patients, 33 (94%) patients launched the app at least once a week, whereas only 2 (6%) patients showed a lower frequency (twice or 5 times in 90 days of observation). The median number of accesses per patient per day was 2 (range 1-30 and IQR 1-3), and approximately 25% (9/35) of the patients accessed the system >3 times a day.

The distribution of app launches according to the time slots is presented in Table 3. Most patients used the app in the morning (2428/5186, 46.81% of launches from 6 to 11:59 AM), an expected result as the mobile app reminded patients to take their medications and enter vital signs mainly at this time. There were no significant differences in the rate of app use related to sex, age, or anticancer treatment (Table 4).

At each trial site, the health care professionals (medical oncologist or nurse) involved in the study accessed the system every 24 to 48 hours and before each study visit to check patient status (data not shown).

Table 3. Time slots of app launches, conversation starts, and vital sign entering.

Time	App launches (n=5186), n (%)	Conversations (n=100), n (%)	Vital signs (n=1757), n (%)
From 6 to 11:59 AM	2428 (46.98)	54 (54)	1085 (61.75)
From noon to 5:59 PM	1143 (22.11)	31 (31)	173 (9.85)
From 6 to 11:59 PM	1524 (29.49)	13 (13)	143 (8.14)
From midnight to 05:59 AM	91 (1.76)	2 (2)	356 (20.26)

Table 4. System usability (app launches, vital sign entering, and messages) according to patient characteristics.

Characteristic	App launches			Vital sign entering			Messages		
	Patients, n (%)	Median days of use/total days, % (IQR)	*P* value (Wilcoxon test)	Patients, n (%)	Median days of use/total days, % (IQR)	*P* value (Wilcoxon test)	None, n (%)	≥1, n (%)	*P* value (Fisher exact test)
Overall patients	35	67 (39-85)	N/A[a]	26	61 (8-100)	N/A	11 (29)	27 (71)	N/A
Gender			.61			.29			.30
Female	19	76 (39-89)		14	54 (9-100)		8 (73)	14 (52)	
Male	16	67 (32-77)		12	61 (6-100)		3 (27)	13 (48)	
Age (years)			.44			.29			.07
<66	16	56 (28-83)		13	95 (9-100)		8 (73)	11 (41)	
≥66	19	74 (40-85)		13	27 (6-90)		3 (27)	16 (59)	
Drug			.33			.32			.29
Capecitabine	18	72 (40-89)		12	10 (4-100)		8 (73)	13 (48)	
TKIs[b]	17	67 (19-80)		14	87 (27-100)		3 (27)	14 (52)	

[a]N/A: not applicable.[b]TKI: tyrosine kinase inhibitor.

Vital Sign Entering

Approximately 68% (26/38) of patients entered ≥1 parameter during the study, the main items being blood pressure (22/26, 85%), pulse rate (9/26, 35%), weight (18/26, 69%), and body temperature (8/26, 31%). Most parameters were entered in the morning (1085/1757, 61.75% from 6 to 11:59 AM), and some were registered in the afternoon (173/1757, 9.84% from noon to 5:59 PM), in the evening (143/1757, 8.14% from 6 to 11:59 PM), and during the night (356/1757, 20.26% from midnight to 5:59 AM), the latter mainly because of patients entering the information in the early hours of the morning (Table 3).

The rate of app use for vital sign entering, calculated as the number of days in which ≥1 parameter was registered with respect to the total number of days of observation, was 61%. There were no significant differences in the rate of vital sign entering related to sex, age, or anticancer treatment. Vital signs were entered more frequently by younger patients (95% <66 years vs 27% ≥66 years) and by patients receiving TKIs (87% vs 10% for capecitabine; Table 4).

Use of System for Messages

Approximately 71% (27/38) of patients used the app for messages at least once during the study, and 212 messages were generated in 100 conversations between patients and health care professionals. The conversations mainly regarded problems with side effects (59/100, 59%), difficulties with the app (30/100, 30%), and clarification about appointments (11/100, 11%). Most messages were sent during the morning hours (54/100, 54% from 6 to 11:59 AM) or in the afternoon (31/100, 31% from noon to 5:59 PM; Table 3). There were no significant differences in the use of the messaging system related to sex, age, or anticancer treatment. However, elderly patients used the messaging system more frequently (Table 4); 59% (16/27) of the patients who sent ≥1 message were aged >65 years, whereas approximately 73% (8/11) of the patients who did not use the messaging system were aged <65 years.

Alarm System

The ONCO-TreC was endowed with a rule-based alarm system [21,22] defined by our oncologists and drawing on state-of-the-art knowledge in the area of cancer. Alarms went off at least once for 23 patients, and a total of 150 alarms were activated. Among the reasons for alarm activation were data not entered for 3 days in a row (22/150, 14.7%), anticancer drugs not taken for 3 days in a row (47/150, 31.3%), systolic blood pressure ≥160 mmHg or diastolic blood pressure ≥100 mmHg (69/150, 46%), and grade ≥3 AEs (12/150, 8%). The median time to resolution by health care professionals was 2 days.

FACT-G and HADS Questionnaires

A total of 36 patients were evaluated by administering questionnaires at baseline, of whom 34 (94%) completed the FACT-G questionnaire and 35 (97%) completed the HADS

questionnaire. Patients showed a mean total score of 69.1 (SD 14.9), reporting better physical well-being (mean 22.5, SD 5.1) but poorer functional well-being (mean 11.4, SD 5.4). Of the 35 cases evaluated by the HADS questionnaire at baseline, 20 (57%) and 19 (54%) were considered to have normal levels of anxiety and depression, respectively.

Data from the FACT-G and HADS questionnaires administered at baseline and at the end of the study were available for only 18 and 20 patients, respectively. In the subgroup of patients who answered both the baseline and posttreatment FACT-G questionnaire, a slight reduction in physical well-being emerged (P=.046), whereas the scores for the other subscales did not change substantially. Of the 20 patients with both pre- and posttreatment HADS, 3 (15%) showed an improved anxiety score after treatment, whereas another 3 (15%) had a worse score (P=.55). Similar results were reported in the depression subscale, with 4 (20%) patients achieving a better status after treatment, and another 4 (20%) showing a worse status (P=.77).

Usability and Acceptability

The patients did not encounter any significant issues in interacting with the app. The SUS scores indicated an excellent subjective assessment of the usability of the system. The overall score was <68 (the average SUS score) in only a small percentage of patients (11.8%); however, it still denoted an acceptable overall usability of the system. An overall score of >80.3 was recorded for 76.4% of patients, which is considered the threshold of excellence [26]. Of note, few patients reported being *worried* about their ability to use the system before they started the trial, and some mentioned having requested the help of a relative the first time they used it. However, all patients reported being able to use ONCO-TreC by themselves within a short period.

The acceptability of ONCO-TreC and its perceived benefits was explored using 2 ad hoc questionnaires distributed to patients at the end of the trial. The overall satisfaction was *very satisfied* (10/20, 50%), *moderately satisfied* (9/20, 45%), *slightly satisfied* (5/20, 5%), and *not satisfied* (0/20, 0%). The patients reported an overall positive effect on several aspects of health care management, measured using a Likert scale. ONCO-TreC proved useful for self-management purposes and was considered a valuable reminder of when to take therapy (19/20, 95% *strongly agreed* and 1/20, 5% *agreed*) and a useful AE tracker (14/20, 70% *strongly agreed* and 5/20, 25% *agreed*). The questionnaire also highlighted a significantly positive effect on the patient–provider relationship as patients felt that ONCO-TreC reinforced the perception of being continuously monitored by providers (18/20,90% *strongly agreed* and 2/20,10% *agreed*). An overall positive evaluation was also recorded for the messaging system as a useful communication tool (12/20, 60% *strongly agreed* and 4/20, 20% *agreed*).

Patient–Provider Communication

ONCO-TreC modified the way in which patients were able to communicate with health care professionals. An ad hoc questionnaire was developed to investigate the communication channels used by patients before and after the introduction of the system. Of the 35 patients who completed the baseline questionnaire, only 20 (57%) returned to the questionnaire at the end of the treatment. Table 5 shows the results of the pre- and posttreatment evaluations for this subgroup. Although the small number of participants and the brevity of the study period did not allow for any definitive conclusions to be drawn about this, our results suggest that the ONCO-TreC messaging system helped reduce all forms of direct contact that would have resulted in a disruption of the workflow in a health care setting (eg, phone calls to a switchboard or physician and physically going to the hospital). Although only 71% (27/38) of the patients used the messaging system integrated into the platform, our findings indicate that it minimized synchronous interactions by favoring asynchronous communication.

Table 5. Channels to contact providers before and after the introduction of ONCO-TreC (N=20).

Contact with the oncology department in the previous couple of months	Often or sometimes, n (%)		Rarely or never, n (%)	
	Before	After	Before	After
Phone call to switchboard	10 (50)	7 (35)	10 (50)	13 (65)
Phone call to physician	6 (30)	4 (20)	14 (70)	16 (80)
Email to physician	4 (20)	2 (10)	16 (80)	18 (90)
Going to the department	13 (65)	7 (35)	7 (35)	13 (65)
ONCO-TreC system	N/A[a]	7 (35)	N/A	13 (65)
Other	1 (5)	0 (0)	19 (95)	20 (100)

[a]N/A: not applicable.

Discussion

Principal Findings

This project aimed to fine-tune and validate an oral anticancer therapy monitoring system in terms of its ability to increase adherence to treatment, improve home management of treatment side effects, and enhance patient or health care professional communication. Adherence is normally considered as the percentage of the prescribed treatment dose actually taken by the patient over a specified period. The most common and simple method used to measure adherence other than direct patient questioning is pill count (ie, counting the number of pills that remain in a patient's medication bottles or vials). Although the simplicity and empirical nature of this method are attractive to many investigators, the method is subject to problems as patients may switch medicines between bottles and discard pills

before visits to appear to be following the regimen. Improving communication between health care professionals and patients through new technologies such as reminders and PDAs might thus be an effective strategy to increase adherence. In this trial, adherence was measured both by pill count and by the number of pills counted by the system (entered every day by the patients at home). Adherence was very high and may have been due, among other factors, to the reminder system integrated into the app, which was considered very useful by virtually all patients. However, the small sample size and absence of a control arm did not allow us to draw definitive conclusions about the efficacy of the system in improving adherence. The ability of the system to measure adherence to treatment was also high, with a concordance of 97.3%.

The reliability of the platform to register toxicity was also investigated by comparing the type and grade of toxicities recorded in the system with those reported by patients during the clinical visits. The quality of the system proved inadequate in that only a fraction of AEs reported by physicians in CRFs were also recorded in the app by patients. This negative result was probably related to some patients not using the app to register toxicity because of technical difficulties, underestimation of the importance of the AE, or other unknown causes. In fact, of the 29 patients reporting AEs during the study visits, only 17 (59%) entered information on the events into the system. Conversely, the patients using the app for AE entering were very meticulous and provided a large amount of data (718 AEs), enabling clinicians to create an accurate, day-to-day reconstruction of the toxicity trend over time. It can thus be hypothesized that, although ONCO-TreC has a high capacity for detecting and monitoring AEs, patients need to be made aware of the importance of this complex part of the system and trained to use it correctly.

The system helped reshape the patient–provider interaction between clinical visits, enabling some synchronous communication to be turned into asynchronous contact through the messaging system. Even though the reshaping of communication was not a primary end point of this study, the decrease in phone calls and patients physically going to the hospital in favor of texting through the platform could make communication more manageable during home oral anticancer treatment. Phone calls and visits tend to disrupt the workflow of hospital departments, whereas asynchronous communication allows providers to make good use of downtime or even define a specific time slot to interact with patients at home. Partial confirmation of the potential usefulness of messaging can be obtained by comparing it with alarm management. In fact, alarms are the counterpart of messages. The former go off automatically when the rule-based system detects a potential harmful pattern, whereas the latter are sent by patients because they need to communicate something. However, in this experimental phase, both alarms and messages addressed noncritical issues.

The use and usability of the system were also analyzed extensively. Approximately 94% (33/35) of patients launched the app at least once a week, and the median number of daily accesses per patient was 2. Most patients used the app for messages and vital sign entering at least once during the study period. The usability of the system was greatly appreciated by patients, and preliminary results of the training step led to the extension of the age limit for enrollment in the trial. This result was probably due to the use of a simplified interface with visualization tools that patients were familiar with and to the work carried out in the preliminary phase of the study with representatives of patient associations. However, as previously mentioned, only some patients used the most advanced functions of the app, that is, messaging and AE reporting. Thus, other studies are warranted to investigate the usability of each function and determine whether their limited use is a result of their design or a lack of interest on the part of the patients.

Conclusions

The ONCO-TreC system appears to be an important tool for improving the home management of and monitoring adherence to oral anticancer drugs. Although the usability and acceptability of the system were very high, its reliability in registering toxicity needs to be improved. A phase III trial comparing the ONCO-TreC system with a standard oral anticancer treatment diary is currently recruiting patients (ClinicalTrials.gov ID NCT04826458).

Acknowledgments

The authors wish to thank Enzo Galligioni for his important contribution to the study design and Gráinne Tierney for editorial assistance. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The project was submitted in 2016 to the Italian Ministry of Health Call for Targeted Research Grant (project code RF-2016-02362926). This study was reviewed and approved by the ethics committee of the Area Vasta Romagna and the Istituto Romagnolo per lo Studio dei Tumori (approval number 1315 of 16/04/2015) and was conducted in accordance with the ethical standards laid down in the 1964 Declaration of Helsinki and later versions. The participants provided written informed consent to participate in this study.

Authors' Contributions

AP, RV, and SF conceived the idea for the study. AP, RV, DA, CE, CAT, AZ, MD, EMP, OC, and SF designed the study. Patient data were collected by DA, FGS, GB, and MD. AP, RV, FF, and EMP were responsible for data interpretation. FF performed the statistical analyses. AP, FF, and EMP drafted the manuscript. All authors were involved in reviewing and commenting on the manuscript and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
App self-reported items.
[DOCX File , 15 KB - jmir_v24i1e27349_app1.docx]

References

1. O'Neill VJ, Twelves CJ. Oral cancer treatment: developments in chemotherapy and beyond. Br J Cancer 2002 Oct 21;87(9):933-937 [FREE Full text] [doi: 10.1038/sj.bjc.6600591] [Medline: 12434279]
2. Findlay M, von Minckwitz G, Wardley A. Effective oral chemotherapy for breast cancer: pillars of strength. Ann Oncol 2008 Feb;19(2):212-222 [FREE Full text] [doi: 10.1093/annonc/mdm285] [Medline: 18006898]
3. Borner M, Scheithauer W, Twelves C, Maroun J, Wilke H. Answering patients' needs: oral alternatives to intravenous therapy. Oncologist 2001;6 Suppl 4:12-16 [FREE Full text] [doi: 10.1634/theoncologist.6-suppl_4-12] [Medline: 11585969]
4. Weingart SN, Brown E, Bach PB, Eng K, Johnson SA, Kuzel TM, et al. NCCN Task Force Report: oral chemotherapy. J Natl Compr Canc Netw 2008 Mar;6 Suppl 3:S1-14. [Medline: 18377852]
5. Osterberg L, Blaschke T. Adherence to medication. N Engl J Med 2005 Aug 04;353(5):487-497. [doi: 10.1056/NEJMra050100] [Medline: 16079372]
6. Partridge AH, Avorn J, Wang PS, Winer EP. Adherence to therapy with oral antineoplastic agents. J Natl Cancer Inst 2002 May 01;94(9):652-661. [doi: 10.1093/jnci/94.9.652] [Medline: 11983753]
7. Mathes T, Pieper D, Antoine S, Eikermann M. Adherence influencing factors in patients taking oral anticancer agents: a systematic review. Cancer Epidemiol 2014 Jun;38(3):214-226. [doi: 10.1016/j.canep.2014.03.012] [Medline: 24768601]
8. Murphy CC, Bartholomew LK, Carpentier MY, Bluethmann SM, Vernon SW. Adherence to adjuvant hormonal therapy among breast cancer survivors in clinical practice: a systematic review. Breast Cancer Res Treat 2012 Jul;134(2):459-478 [FREE Full text] [doi: 10.1007/s10549-012-2114-5] [Medline: 22689091]
9. Al-Barrak J, Cheung WY. Adherence to imatinib therapy in gastrointestinal stromal tumors and chronic myeloid leukemia. Support Care Cancer 2013 Aug;21(8):2351-2357. [doi: 10.1007/s00520-013-1831-6] [Medline: 23708821]
10. Given BA, Spoelstra SL, Grant M. The challenges of oral agents as antineoplastic treatments. Semin Oncol Nurs 2011 May;27(2):93-103. [doi: 10.1016/j.soncn.2011.02.003] [Medline: 21514479]
11. Makubate B, Donnan PT, Dewar JA, Thompson AM, McCowan C. Cohort study of adherence to adjuvant endocrine therapy, breast cancer recurrence and mortality. Br J Cancer 2013 Apr 16;108(7):1515-1524 [FREE Full text] [doi: 10.1038/bjc.2013.116] [Medline: 23519057]
12. Wu EQ, Johnson S, Beaulieu N, Arana M, Bollu V, Guo A, et al. Healthcare resource utilization and costs associated with non-adherence to imatinib treatment in chronic myeloid leukemia patients. Curr Med Res Opin 2010 Jan;26(1):61-69. [doi: 10.1185/03007990903396469] [Medline: 19905880]
13. Ganesan P, Sagar TG, Dubashi B, Rajendranath R, Kannan K, Cyriac S, et al. Nonadherence to imatinib adversely affects event free survival in chronic phase chronic myeloid leukemia. Am J Hematol 2011 Jun;86(6):471-474 [FREE Full text] [doi: 10.1002/ajh.22019] [Medline: 21538468]
14. Hershman DL, Shao T, Kushi LH, Buono D, Tsai WY, Fehrenbacher L, et al. Early discontinuation and non-adherence to adjuvant hormonal therapy are associated with increased mortality in women with breast cancer. Breast Cancer Res Treat 2011 Apr;126(2):529-537 [FREE Full text] [doi: 10.1007/s10549-010-1132-4] [Medline: 20803066]
15. McCowan C, Wang S, Thompson AM, Makubate B, Petrie DJ. The value of high adherence to tamoxifen in women with breast cancer: a community-based cohort study. Br J Cancer 2013 Sep 03;109(5):1172-1180 [FREE Full text] [doi: 10.1038/bjc.2013.464] [Medline: 23949153]
16. Jackevicius CA, Mamdani M, Tu JV. Adherence with statin therapy in elderly patients with and without acute coronary syndromes. JAMA 2002;288(4):462-467. [doi: 10.1001/jama.288.4.462] [Medline: 12132976]
17. Cramer J, Rosenheck R, Kirk G, Krol W, Krystal J, VA Naltrexone Study Group 425. Medication compliance feedback and monitoring in a clinical trial: predictors and outcomes. Value Health 2003;6(5):566-573 [FREE Full text] [doi: 10.1046/j.1524-4733.2003.65269.x] [Medline: 14627063]
18. Haynes RB, McDonald HP, Garg AX. Helping patients follow prescribed treatment: clinical applications. JAMA 2002 Dec 11;288(22):2880-2883. [doi: 10.1001/jama.288.22.2880] [Medline: 12472330]
19. Patton K, Meyers J, Lewis BE. Enhancement of compliance among patients with hypertension. Am J Manag Care 1997 Nov;3(11):1693-1698 [FREE Full text] [Medline: 10178467]
20. Ran M, Xiang M, Chan CL, Leff J, Simpson P, Huang M, et al. Effectiveness of psychoeducational intervention for rural Chinese families experiencing schizophrenia--a randomised controlled trial. Soc Psychiatry Psychiatr Epidemiol 2003 Feb;38(2):69-75. [doi: 10.1007/s00127-003-0601-z] [Medline: 12563548]

21. Galligioni E, Piras EM, Galvagni M, Eccher C, Caramatti S, Zanolli D, et al. Integrating mHealth in oncology: experience in the province of Trento. J Med Internet Res 2015;17(5):e114 [FREE Full text] [doi: 10.2196/jmir.3743] [Medline: 25972226]
22. Passardi A, Rizzo M, Maines F, Tondini C, Zambelli A, Vespignani R, et al. Optimisation and validation of a remote monitoring system (Onco-TreC) for home-based management of oral anticancer therapies: an Italian multicentre feasibility study. BMJ Open 2017 May 29;7(5):e014617 [FREE Full text] [doi: 10.1136/bmjopen-2016-014617] [Medline: 28554917]
23. FACIT Homepage. FACT-G. URL: http://www.facit.org/ [accessed 2021-12-08]
24. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
25. Brooke J. SUS: A "quick and dirty" usability scale. In: Usability Evaluation in Industry. Boca Raton, Florida: CRC Press; 1996.
26. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 2009 May;4(3):114-123. [doi: 10.5555/2835587.2835589]

Abbreviations

AE: adverse event
CRF: case report form
FACT-G: Functional Assessment of Cancer Therapy-General
HADS: Hospital Anxiety and Depression Scale
SUS: System Usability Scale
TKI: tyrosine kinase inhibitor

Edited by R Kukafka; submitted 22.01.21; peer-reviewed by S Davis, S Nabhani-Gebara, M Lozano-Lozano, V Shih; comments to author 09.04.21; revised version received 10.06.21; accepted 10.11.21; published 26.01.22.

Please cite as:
Passardi A, Foca F, Caffo O, Tondini CA, Zambelli A, Vespignani R, Bartolini G, Sullo FG, Andreis D, Dianti M, Eccher C, Piras EM, Forti S
A Remote Monitoring System to Optimize the Home Management of Oral Anticancer Therapies (ONCO-TreC): Prospective Training–Validation Trial
J Med Internet Res 2022;24(1):e27349
URL: https://www.jmir.org/2022/1/e27349
doi:10.2196/27349
PMID:35080505

Original Paper

The Impact of Synchronous Telehealth Services With a Digital Platform on Day-by-Day Home Blood Pressure Variability in Patients with Cardiovascular Diseases: Retrospective Cohort Study

Ying-Hsien Chen[1], MD; Chi-Sheng Hung[1], MD, PhD; Ching-Chang Huang[1], MD; Jen-Kuang Lee[1], MD, PhD; Jiun-Yu Yu[2], PhD; Yi-Lwun Ho[1], MD, PhD

[1]Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan

[2]Department of Business Administration, College of Management, National Taiwan University, Taipei, Taiwan

Corresponding Author:
Yi-Lwun Ho, MD, PhD
Department of Internal Medicine
National Taiwan University Hospital
#7, Chung-Shan South Road
Taipei, 100225
Taiwan
Phone: 886 2 23123456 ext 65524
Fax: 886 2 2353 9346
Email: ylho@ntu.edu.tw

Abstract

Background: Hypertension is associated with a large global disease burden with variable control rates across different regions and races. Telehealth has recently emerged as a health care strategy for managing chronic diseases, but there are few reports regarding the effects of synchronous telehealth services on home blood pressure (BP) control and variability.

Objective: The objective of this study is to investigate the effect of synchronous telehealth services with a digital platform on home BP.

Methods: This retrospective study was conducted by the Taiwan ELEctroHEALTH study group at the Telehealth Center of the National Taiwan University Hospital. We analyzed home BP data taken from 2888 patients with cardiovascular disease (CVD) enrolled in our telehealth program between 2009 to 2017. Of the 2888 patients with CVD, 348 (12.05%) patients who received home BP surveillance for ≥56 days were selected for BP analysis. Patients were stratified into three groups: (1) poorly controlled hypertension, (2) well-controlled hypertension, and (3) nonhypertension. The mean, SD, coefficient of variation (CV), and average real variability were calculated.

Results: Telehealth interventions significantly and steadily reduced systolic blood pressure (SBP) in the poorly controlled hypertension group from 144.8.2±9.2 to 133.7±10.2 mmHg after 2 months (P<.001). BP variability reduced in all patients: SBP-SD decreased from 7.8±3.4 to 7.3±3.4 after 2 months (P=.004), and SBP-CV decreased from 6.3±2.5 to 5.9±2.6 after 2 months (P=.004). Event-free survival (admission) analysis stratified by SBP-SD showed longer time to first hospitalization for Q1 patients compared with Q4 patients (P=.02, odds ratio 2.15, 95% CI 1.18-3.89).

Conclusions: Synchronous telehealth intervention may improve home BP control and decrease day-by-day home BP variability in patients with CVD.

(J Med Internet Res 2022;24(1):e22957) doi:10.2196/22957

KEYWORDS

blood pressure; variability; telehealth; hypertension; cardiovascular disease; chronic disease; heart; digital platform; cohort; management; intervention

Introduction

According to data sourced from the noncommunicable disease (NCD) Risk Factor Collaboration, the number of adults worldwide with increased blood pressure (BP) was 1.13 billion in 2015, up from 594 million in 1975 [1]. As the number of adults with elevated BP levels continues to rise, hypertension is becoming a growing public health threat and an important global NCD. Globally, hypertension is considered a leading risk factor for both cardiovascular disease (CVD) and chronic kidney disease [2]. The relationship between the risks of CVD and BP can be continuous or log-linear [3]. Starting at BP levels of 115/75 mmHg, mortality risks associated with stroke, heart disease, or other CVDs are doubled with every 20 mmHg increase in systolic blood pressure (SBP) and every 10 mmHg increase in diastolic blood pressure (DBP) [3], while treatment to control BP can provide proportional reductions in risk that are independent of pretreatment BP levels [4].

The best hypertension management practices combine pharmacotherapy and nonpharmacological interventions, such as weight reduction, increased physical activity, the Dietary Approaches to Stop Hypertension (DASH) diet, potassium supplements, decreased sodium intake, and reduced alcohol consumption [5-7]. Despite widespread initiatives for treating hypertension and the availability of antihypertensive medications, more than 50% people in the United States have uncontrolled, elevated BP [8]. Among adults with hypertension in the United States, hypertension control rates are around 50%-60% for non-Hispanic Whites but less than 50% for non-Hispanic Black, Hispanic, and non-Hispanic Asian populations [9]. This phenomenon is primarily attributed to a lack of awareness and treatment, as well as suboptimal adherence to self-care [9,10]. It is recommended that structured and team-based interventions be paired with technology-based strategies rooted in health information to control and treat hypertension [6]. Telehealth strategies, such as telemedicine, eHealth, and mHealth, are innovative tools for BP control. Meta-analyses of randomized controlled trials (RCTs) using different telehealth interventions have demonstrated greater BP reductions and a larger proportion of patients achieving BP control [11,12]. However, the involvement of various telehealth interventions within a single strategy can lead to inconsistent clinical effects. Anker et al [13] proposed a classification system enabling comparison between various telehealth strategies for the four different generations of telemedicine based on data transfer procedures, the analytical ability of the telehealth system, and the degree of integration with patient primary health care structures. The telehealth service from the National Taiwan University Hospital (NTUH), Taipei, Taiwan, has provided a fully integrated remote management system with constant analytical and decision-making support for full therapeutic authority outside office hours since 2009. This service model fulfills the criteria for the fourth-generation telehealth service/program proposed by Anker et al [13]. As evidence from use of a fourth-generation telehealth program for home BP control and home BP variability is currently lacking, we decided to analyze patient data collected through the Telehealth Center of the NTUH to determine the impacts of our telehealth intervention services on home BP control and home BP variability. The objective of our study was to investigate the effect of synchronous telehealth services with a digital platform on home BP.

Methods

Study Design

This was a single-center retrospective cohort study approved by the institutional review board of the NTUH and conducted by the Taiwan ELEctroHEALTH (TELEHEALTH) study group. We obtained written informed consent from all participants for telehealth services.

Telehealth Services

Since 2009, the Telehealth Center of the NTUH has provided telehealth services for patients with CVD, including both patients with multiple CVD risk factors and patients with established CVDs, such as coronary artery disease, myocardial infarction, congestive heart failure, arrhythmia, and other surgical or congenital heart conditions. Detailed descriptions of enrollment criteria and the scope of telehealth services have been reported previously [14].

Our telehealth services encompassed surveillance of biometric data, discussions between patients and the Telehealth Center, and full therapeutic authority and suggestions from case managers and cardiologists 24 hours a day. Patients with CVD were required to measure vital biometric data (single-lead electrocardiography, BP, pulse rate [PR], finger-stick glucose, and oxygen saturation) at home daily or on demand. Data were then transmitted to a cloud database for review and evaluation by nurse case managers or physicians who were able to access the data via an internet-based interactive platform and provide instant advice according to the clinical conditions of each patient [14-17]. The internet-based interactive platform used by the Telehealth Center was developed by the Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taiwan [18]. The ultimate goal of the telehealth program was to bridge the gap between acute care and home care and also to emphasize education, enhance self-awareness, improve self-care skills, strengthen prevention, and facilitate early detection of clinical deterioration.

Patient Selection

To evaluate day-by-day BP changes over a period of 8 weeks, patients with CVD who had received telehealth services and recorded daily home BP measurements consecutively for more than 56 days (8 weeks) during 2009 to 2017 were retrospectively selected for analysis. Patients who were enrolled in the telehealth program for less than 56 days or those who failed to transmit daily home BP measurements were excluded from our analysis. No statistical power calculation was conducted prior to the study. The sample size was based on the available data. All statistical tests were determined after examination of available data.

Home BP

We compiled SBP, DBP, and PR data from all patients to serve as explanatory variables. Mean blood pressure (MBP) values

were calculated using the formula MBP = [SBP + (2 × DBP)]/3. We calculated the daily home MBP values for every selected patient from days 1 to 56 and then calculated baseline home BP by averaging the first 3 days (days 1-3) of home BP values. Thereafter, average posttelehealth BP values were calculated in 14-day (2-week) intervals: values for week 2 were calculated by averaging BP values from days 4 to 14, values for week 4 were calculated by averaging BP values from days 15 to 28, values for week 6 were calculated by averaging BP values from days 29 to 42, and values for week 8 were calculated by averaging BP values from days 43 to 56. Calculations were conducted in the same manner for SBP, DBP, MBP, and PR data.

Home BP Variability

Home BP variability values were calculated by SD, coefficient of variation (CV), and average real variability (ARV) [19]. Home BP variability was calculated in 14-day (2-week) intervals: baseline values were derived from days 1 to 14, weeks 3-4 values from days 15 to 28, weeks 5-6 values from days 29 to 42, and weeks 7-8 values from days 43 to 56. Calculations were conducted in the same manner for SBP and DBP data.

Group Analysis

We divided patients into the following groups based on clinical diagnosis and BP data at baseline: (1) poorly controlled hypertension group, which included hypertension patients with baseline home SBP ≥135 mmHg or home DBP ≥85 mmHg according to European Society of Cardiology (ESC) guideline definitions [7]; (2) well-controlled hypertension group, which included hypertension patients with baseline home SBP <135 mmHg and home DBP <85 mmHg; and (3) nonhypertension group, which included patients who had not previously been diagnosed with hypertension. All group analyses were determined based on examinations of available data.

Outcome Measurement and Follow-Up

Mean values and variability indicators for home BP levels after telehealth services were compared with baseline values (at prespecified durations after telehealth services) at 2-week intervals. Clinical outcomes regarding the time to first hospitalization were censored in selected patients with CVD with a mean follow-up period of 22 months.

Statistical Analysis

Continuous data are presented as the mean±SD. Discrete data are given as counts and percentages. Paired *t*-tests were used to analyze changes in different data sets over time for the same individuals compared with baseline. Between-group comparisons of discrete indicators at baseline (gender proportions, comorbidities, and medications) were analyzed using a chi-square test, while between-group comparisons of continuous indicators at baseline (age, average BP measurements per day, hemoglobin A_{1c} (HbA_{1c}) levels, creatinine levels, and total cholesterol) were calculated using an ANOVA test.

We used paired *t*-tests to compare differences in BP values between patient groups. Basic assumptions for paired *t*-tests were verified prior to conducting these analyses. Independently observed continuous BP values were set as our dependent variable. SBP values <65 mmHg or >250 mmHg and DBP values <40 mmHg or >150 mmHg were excluded to avoid outliers. Two-week average BP values were found to be normally distributed based on graphical presentations of QQ plots and histograms. One-way ANOVA and chi-square tests were used to compare differences in baseline characteristics between patient groups. Basic assumptions for one-way ANOVA and chi-square tests were verified prior to conducting these analyses. One-way ANOVA was used for continuous data, such as age, average BP measurements per day, HbA_{1c}, creatinine, and total cholesterol. All data adhered to ANOVA assumptions of normality (based on QQ plots), homoscedasticity (based on scatterplot of residuals), no multicollinearity (based on correlation data and variance inflation factor), and independence of measurements. Chi-square tests were used for independent categorical data such as gender, comorbidities, and medication.

A two-sided *P* value of <0.05 was considered statistically significant. Kaplan-Meier analysis was used to compare time to first hospitalization for different groups, and the log-rank test (pairwise over strata) was used to detect differences between survival curves. Patients were censored if they were not hospitalized more than 36 months after entry into our telehealth program or if there were insufficient data to indicate that they were hospitalized during this period. IBM SPSS for Windows, version 22.0, (SPSS Inc., Chicago, IL, USA) was used for statistical analyses.

Results

Descriptive Statistics

From 2009 to 2017, a total of 2888 patients with CVD received synchronous telehealth services and delivered a total of 678,596 BP data sets over varying lengths of time. Of these, 2540 patients had insufficient data over a 56-day period and were excluded from the study. The majority of missing data in the cohort was due to participants forgetting to take BP measurements or leaving the program before 56 days. A total of 348 patients (mean age 67.6±14.6 years; 232 [66.7%] males) who recorded daily home BP measurements for more than 56 days were selected for BP analysis. During their 56-day period of telehealth care (days 1-56), 383,400 BP data sets were transmitted for telehealth surveillance.

The patients were further stratified into the (1) poorly controlled hypertension group (n=81), (2) well-controlled hypertension group (n=125), and (3) nonhypertension group (n=142) according to prespecified definitions. Baseline demographics and medications before telehealth care are summarized in Table 1. In the poorly controlled hypertension group, the mean age was significantly higher (*P*=.001) and patients had more comorbidities and higher rates of renin-angiotensin system (RAS) blocker usage compared with the other groups.

Table 1. Baseline characteristics of telehealth service patients.

Characteristics, diseases, and medications	Poorly controlled hypertension	Well-controlled hypertension	Nonhypertension	Total patients	*P* value
Characteristics					
Patients	n=81	n=125	n=142	N=348	—[a]
Age (years), mean±SD	72.0±12.2	68.1±13.8	64.6±15.8	67.6±14.6	.001
Male, n (%)	52 (64%)	79 (63.2%)	101 (71.1%)	232 (66.7%)	.34
Diseases					
Hypertension, n (%)	81 (100%)	125 (100%)	0 (0.0%)	183 (52.6%)	<.001
Heart failure, n (%)	15 (19%)	29 (23.2%)	39 (27.5%)	83 (23.9%)	.31
Diabetes mellitus, n (%)	37 (46%)	55 (44.0%)	18 (12.7%)	110 (31.6%)	<.001
Hyperlipidemia, n (%)	24 (30%)	64 (51.2%)	30 (21.1%)	118 (33.9%)	<.001
Myocardial infarction, n (%)	9 (11%)	17 (13.6%)	18 (12.7%)	44 (12.6%)	.87
Coronary artery disease, n (%)	37 (46%)	72 (57.6%)	57 (40.1%)	166 (47.7%)	.02
Stroke, n (%)	6 (7%)	23 (18.4%)	6 (4.2%)	35 (10.1%)	<.001
End-stage renal failure, n (%)	16 (20%)	14 (11.2%)	4 (2.8%)	34 (9.8%)	<.001
Atrial fibrillation, n (%)	10 (12%)	24 (19.2%)	26 (18.3%)	60 (17.2%)	.40
Average BP[b] measurements per day	2.3±0.9	2.2±1.1	2.1±0.7	2.2±0.9	.21
HbA_{1c}[c] (%), mean±SD	6.5±1.0	6.4±1.0	6.2±1.1	6.3±1.0	.05
Creatinine[d] (mg/dL), mean±SD	1.8±2.3	1.3±1.4	1.3±1.2	1.4±1.5	.06
Total cholesterol (mg/dL), mean±SD	166.6±43.9	158.9±34.8	159.6±36.9	160.9±37.9	.33
Medications					
Calcium channel blocker, n (%)	4 (5%)	12 (9.6%)	8 (5.6%)	24 (6.9%)	.32
Diuretics, n (%)	18 (22%)	33 (26.4%)	24 (16.9%)	75 (21.6%)	.17
Beta blocker, n (%)	16 (20%)	29 (23.2%)	32 (22.5%)	77 (22.1%)	.84
RAS[e] blocker, n (%)	25 (31%)	48 (38.4%)	23 (16.2%)	96 (27.6%)	.001

[a]—: not applicable.

[b]BP: blood pressure.

[c]HbA_{1c}: hemoglobin A_{1c}.

[d]Patients with renal failure under dialysis were excluded.

[e]RAS: renin-angiotensin system. The RAS blocker includes angiotensin-converting enzyme inhibitors, angiotensin receptor blockers, and angiotensin receptor neprilysin inhibitor.

Home BP Change After Telehealth Service

In the poorly controlled hypertension group, BP significantly and constantly decreased after 2 weeks of telehealth service. During the 8 weeks of telehealth service, the mean SBP decreased from baseline 144.8±9.2 mmHg to 133.7±10.2 mmHg (*P*<.001) and the mean DBP also decreased from baseline 77.7±12.7 mmHg to 72.9±12.1 mmHg (*P*<.001).

In the well-controlled hypertension group, a gradual increment in BP was observed. During the 8 weeks of telehealth service, the mean SBP increased from baseline 119.7±9.8 mmHg to 124.1±11.3 mmHg (*P*<.001) and the mean DBP increased from baseline 71.3±9.9 mmHg to 73.0±9.4 mmHg (*P*=.009). A slight increment in home BP (1.7 mmHg in SBP and 1.6 mmHg in DBP) was also observed in the nonhypertension group during the 8 weeks of telehealth service (Table 2). The effect of telehealth services on the PR and MBP in different subgroups is provided in Multimedia Appendix 1. The impact of telehealth services on BP values in all patients is provided in Multimedia Appendix 2.

Table 2. Change in BP[a] during the period of telehealth service by subgroup.

BP measures	Baseline (days 1-3)	Week 2 (days 4-14)	Week 4 (days 15-28)	Week 6 (days 29-42)	Week 8 (days 43-56)
Poorly controlled hypertension group					
SBP[b] (mmHg), mean±SD	144.8±9.2	140.7±10.0	137.1±12.0	136.0±11.8	133.7±10.2
P value[c]	—[d]	<.001	<.001	<.001	<.001
DBP[e], mmHg	77.7±12.7	76.0±11.0	74.8±12.2	73.9±11.9	72.9±12.1
P value	—	.01	.002	<.001	<.001
Well-controlled hypertension group					
SBP (mmHg), mean±SD	119.7±9.8	121.9±10.6	123.1±10.7	123.3±11.0	124.1±11.3
P value	—	.004	<.001	<.001	<.001
DBP (mmHg), mean±SD	71.3±9.9	72.0±9.4	72.8±9.5	72.8±9.2	73.0±9.4
P value	—	.11	.006	.02	.009
Nonhypertension group					
SBP (mmHg), mean±SD	115.6±11.0	115.5±11.7	116.0±11.6	116.8±11.5	117.3±11.6
P value	—	.81	.56	.13	.04
DBP (mmHg), mean±SD	68.6±8.8	69.1±8.4	69.1±9.1	70.0±9.0	70.2±9.4
P value	—	.25	.35	.02	.01

[a]BP: blood pressure.

[b]SBP: systolic blood pressure.

[c]Data were expressed as the mean±SD and were compared with baseline.

[d]—: not applicable.

[e]DBP: diastolic blood pressure.

Change in Home BP Variability After Telehealth Service

Home day-by-day BP variability in different subgroups decreased following telehealth services (Table 3). In the poorly controlled hypertension group, SBP-SD decreased from 9.2±4.1 mmHg (baseline) to 8.2±3.5 mmHg during weeks 5-6 (*P*=.04) and stayed constant thereafter (variability for weeks 7-8 did not rise significantly compared with variability for weeks 5-6). DBP-SD decreased from 5.7±2.5 mmHg (baseline) to 4.9±1.8 mmHg during weeks 5-6 (*P*=.02) and remained stable thereafter.

We found no significant changes with regard to SBP or DBP variability measured by the CV or ARV.

In the well-controlled hypertension group, SBP-CV during weeks 7-8 was 6.0%±2.6%, which was significantly lower than baseline SBP-CV (6.5%±2.3%, *P*=.02), and DBP-CV during weeks 5-6 was 6.9%±3.0%, which was also significantly lower than baseline DBP-CV (7.5%±2.9%, *P*=.01).

In the nonhypertension group, a significant decrease in DBP-SD was observed during weeks 3-4 compared with baseline.

Table 3. Change in BP[a] variability during the period of telehealth service by subgroup.

BP measures	Baseline (days 1-14)	Weeks 3-4 (days 15-28)	Weeks 5-6 (days 29-42)	Weeks 7-8 (days 43-56)
Poorly controlled hypertension group				
SBP[b]-SD (mmHg), mean±SD	9.2±4.1	8.5±3.5	8.2±3.5	8.3±4.1
P value[c]	—[d]	.14	.04	.09
SBP-CV[e] (%), mean±SD	6.5±2.8	6.2±2.6	6.0±2.5	6.2±3.0
P value	—	.39	.16	.40
SBP-ARV[f] (mmHg), mean±SD	8.6±4.0	8.7±3.9	8.6±4.1	8.7±4.0
P value	—	.84	.86	.88
DBP[g]-SD (mmHg), mean±SD	5.7±2.5	5.3±3.3	4.9±1.8	5.0±2.1
P value	—	.43	.02	.06
DBP-CV (%), mean±SD	7.4±3.1	7.3±4.8	6.8±2.7	7.0±2.9
P value	—	.81	.11	.28
DBP-ARV (mmHg), mean±SD	5.3±2.5	5.5±3.1	5.1±1.9	5.3±2.3
P value	—	.55	.49	.98
Well-controlled hypertension group				
SBP-SD (mmHg), mean±SD	8.0±3.2	7.7±3.7	7.5±3.0	7.5±3.3
P value	—	.46	.15	.08
SBP-CV (%), mean±SD	6.5±2.3	6.2±2.8	6.1±2.3	6.0±2.6
P value	—	.28	.08	.02
SBP-ARV (mmHg), mean±SD	8.0±3.5	7.9±3.2	8.0±3.5	8.1±3.6
P value	—	.75	.90	.70
DBP-SD (mmHg), mean±SD	5.3±1.9	5.3±2.4	5.0±2.1	5.3±3.1
P value	—	.99	.05	.99
DBP-CV (%), mean±SD	7.5±2.9	7.3±3.2	6.9±3.0	7.4±4.8
P value	—	.50	.01	.83
DBP-ARV (mmHg), mean±SD	5.3±2.2	5.4±2.5	5.2±2.3	5.8±3.2
P value	—	.67	.69	.08
Nonhypertension group				
SBP-SD (mmHg), mean±SD	6.9±2.9	6.6±2.9	6.7±3.1	6.6±2.8
P value	—	.17	.29	.15
SBP-CV (%), mean±SD	6.0±2.4	5.7±2.3	5.7±2.6	5.6±2.3
P value	—	.11	.19	.08
SBP-ARV (mmHg), mean±SD	6.7±2.9	7.0±3.2	7.2±3.1	7.0±3.3
P value	—	.23	.09	.14
DBP-SD (mmHg), mean±SD	4.9±2.2	4.6±2.0	4.7±2.4	4.9±2.3
P value	—	.05	.27	.96
DBP-CV (%), mean±SD	7.2±3.2	6.7±3.0	6.7±3.3	7.2±3.6
P value	—	.05	.13	.88
DBP-ARV (mmHg), mean±SD	4.8±2.2	4.8±2.1	5.0±2.3	5.1±2.4
P value	—	.96	.56	.19

[a]BP: blood pressure.

[b]SBP: systolic blood pressure.

[c]Data were expressed as the mean±SD and were compared with baseline.
[d]—: not applicable.
[e]CV: coefficient of variation.
[f]ARV: average real variability.
[g]DBP: diastolic blood pressure.

The effect of telehealth services on day-by-day home BP variability in all patients is provided in Table 4. Telehealth services significantly improved SBP variability (SD and CV) during weeks 3-4 and remained constant during weeks 5-6 and weeks 7-8. DBP variability (SD and CV) was also significantly improved at weeks 5-6 compared with baseline.

Table 4. Change in BP[a] variability during the period of telehealth service in all patients (N=348).

BP measures	Baseline (days 1-14)	Weeks 3-4 (days 15-28)	Weeks 5-6 (days 29-42)	Weeks 7-8 (days 43-56)
SBP[b]-SD, mmHg	7.8±3.4	7.5±3.4	7.3±3.2	7.3±3.4
P value[c]	—[d]	.04	.008	.004
SBP-CV[e], %	6.3±2.5	6.0±2.6	5.9±2.5	5.9±2.6
P value	—	.04	.009	.004
SBP-ARV[f], mmHg	7.6±3.5	7.7±3.4	7.8±3.5	7.8±3.6
P value	—	.54	.38	.35
DBP[g]-SD, mmHg	5.2±2.2	5.0±2.5	4.8±2.2	5.1±2.6
P value	—	.11	.003	.31
DBP-CV, %	7.4±3.1	7.1±3.6	6.8±3.1	7.2±4.0
P value	—	.12	.002	.46
DBP-ARV, mmHg	5.1±2.3	5.2±2.5	5.1±2.2	5.4±2.7
P value	—	.55	.85	.06

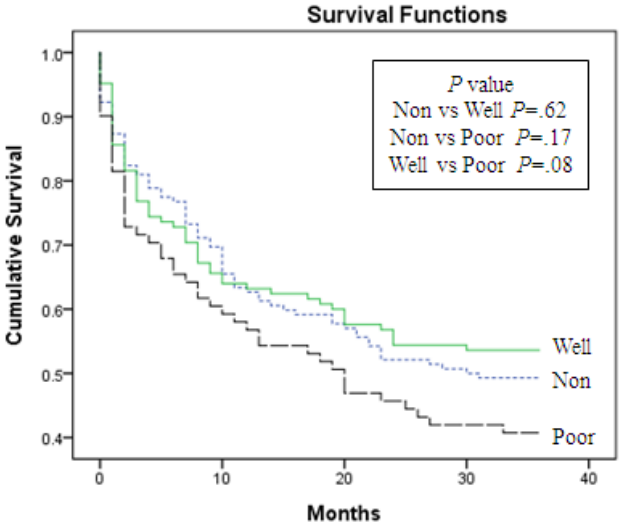
[a]BP: blood pressure.
[b]SBP: systolic blood pressure.
[c]Data were expressed as the mean±SD and were compared with baseline.
[d]—: not applicable.
[e]CV: coefficient of variation.
[f]ARV: average real variability.
[g]DBP: diastolic blood pressure.

Time to First Hospitalization After Telehealth Services

No significant differences were found between the three subgroups in terms of time to first hospitalization (*P*=.20; Figure 1) during a mean follow-up period of 22 months.

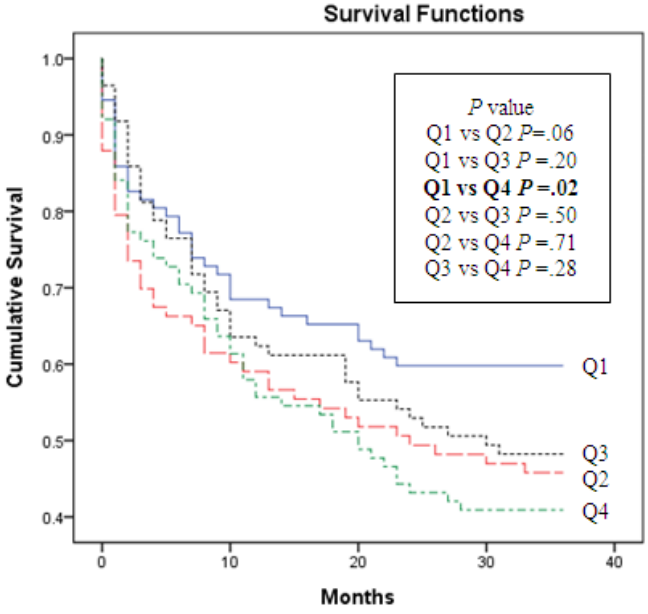
We further stratified these patients according to baseline SBP-SD quartiles (Q1: <5.5; Q2: 5.5-7.1; Q3: 7.1-9.7; and Q4: >9.7), and a significant difference in time to first hospitalization was observed between Q1 and Q4 SBP-SD variability groups (*P*=.02, odds ratio 2.15, 95% CI 1.18-3.89; Figure 2).

Figure 1. Kaplan-Meier survival estimates for time to first hospitalization stratified by hypertension type (poor: poorly controlled hypertension; well: well-controlled hypertension; non: nonhypertension).



	Month 0	Month 10	Month 20	Month 30
Non	142	93	81	71
Well	125	80	72	67
Poor	81	48	38	34

Figure 2. Kaplan-Meier survival estimates for time to first hospitalization across quartiles of day-by-day variability, SD of systolic blood pressure at baseline. Q1 to Q4 indicate ascending quartiles: (Q1: <5.5, Q2: 5.5-7.1, Q3: 7.1-9.7, Q4: >9.7).



	Month 0	Month 10	Month 20	Month 30
Q1	92	63	58	55
Q2	83	50	43	39
Q3	85	54	47	42
Q4	88	54	43	36

Discussion

Principal Findings

The findings from this retrospective study suggest that synchronous telehealth intervention may improve home BP control and decrease home BP variability in patients with CVD.

For patients with poorly controlled hypertension, our telehealth intervention demonstrated a significant BP reduction after only 2 weeks of service. BP reductions were continuous and remained steady throughout the 8-week period of telehealth services, achieving an absolute 11.1 mmHg SBP reduction and a 4.8 mmHg DBP reduction after 8 weeks of telehealth intervention. For patients without hypertension or with well-controlled hypertension, the telehealth intervention helped to normalize home BP, possibly by avoiding overtreatment and hypotension. The telehealth intervention also decreased BP variability after 4 weeks of telehealth services, and this effect remained constant hereafter and may be observed across the three different subgroups.

Telehealth Services Improved Home BP Control

A higher SBP and DBP are known to be associated with increased risk of CVD [20], and high BP variability is associated with cardiovascular end-organ damage, renal outcomes, and increased risk of incident CVD [19,21,22]. Our study therefore demonstrates the value of telehealth services in controlling both home BP and home BP variability, which can potentially help to prevent future occurrence of CVD events.

Our previous study found that all-cause admission rates and all-cause hospital stay durations of patients with CVD decrease following synchronous telehealth interventions [14]. When compared with usual care, synchronous telehealth interventions were found to result in better cost-effectiveness [15] and were also proven to be associated with lower all-cause mortality [16]. It is speculated that these proven clinical benefits stemmed from continuous improvement in the control of chronic conditions, including BP control, which in turn led to reduced mortality. The results from this retrospective study provide additional evidence to support this speculation.

Telehealth Services Normalized Home BP

During the 8 weeks of continuous telehealth intervention, home BP values significantly increased in the well-controlled hypertension group (mean SBP increased from 119.7±9.8 to 124.1±11.3 mmHg) and the nonhypertension group (SBP increased from 115.6±11.0 to 117.3±11.6 mmHg). This paradoxical increase in home BP should be considered a normalization of BP via telehealth interventions, which avoid overtreatment and hypotension. Patients in the nonhypertension group still received calcium channel blockers, beta blockers, diuretics, or RAS blockers for clinical indications other than hypertension, such as arrhythmia or heart failure; this patient population was considered highly susceptible to hypotension during treatment, and it may be that our telehealth interventions helped to prevent hypotensive events. This is worth noting, as a lower BP is associated with a worse prognosis for patients with heart failure [23].

The Systolic Blood Pressure Intervention Trial (SPRINT) study [24] aimed for an SBP goal of <120 mmHg. In the intensive treatment group, a marked decrease in the SBP was accompanied by a marked increase in the incidence of side effects, such as hypotension, syncope, electrolyte abnormalities, and kidney injures; similar results were also observed in the Action to Control Cardiovascular Risk in Diabetes (ACCORD) study [25].

Among treated hypertensive patients, SBP values of <110 mmHg are associated with potential serious falls and syncope [26]. Additionally, hypotension or masked hypotension occurs frequently in older hypertensive adults undergoing treatment [27,28], and low daytime SBP values are also known to be associated with a greater level of cognitive decline in these older patients [29]. Hypotension in hypertensive elderly patients undergoing treatment is an important red flag that makes it unclear whether these patients actually benefit from BP control. Out-of-office BP (including ambulatory BP monitoring or home BP monitoring) is specifically recommended as a method for identifying hypotension side effects during treatment [7]. Synchronous telehealth interventions not only help to identify hypotension side effects but may also exert an additional role in avoiding overtreatment and preventing hypotension.

Other Telehealth Interventions for BP

The American Heart Association (AHA) recommends the use of out-of-office BP measurements paired with telehealth counseling and titration of BP-lowering medication to assess treatment response, improve adherence, and enhance BP control [6].

Meta-analyses of RCTs have shown that highly heterogeneous telehealth intervention models ranging from computer-based support systems to programs led by medical staff can improve BP control and achieve BP normalization in a larger proportion of patients [11,12]. Omboni et al [11] analyzed 23 RCTs (7037 patients) and found that telemonitoring of home BP demonstrates an improvement over office SBP by 4.7 mmHg and office DBP by 2.5 mmHg compared with usual care [11]. The Telemonitoring and/or Self-Monitoring of Blood Pressure in Hypertension (TASMINH4) trial, an unmasked RCT study that compared self-monitored BP with and without telemonitoring in patients with poorly controlled BP, failed to demonstrate any BP control benefits associated with telemonitoring. However, the study was conducted in 142 general practices (GPs) throughout the United Kingdom, which provided a simple and free SMS text-based telemonitoring service. Participants in the telemonitoring group received an SMS message on a weekly basis containing a report of MBP values. Overly high or low readings triggered an alert, which was sent to participants but not GPs, and attending clinicians reviewed the readings on a monthly basis [30]. This asynchronous telemonitoring design meant that medical advice was often delayed for several weeks and was perhaps why the results of the study failed to find benefits associated with telemonitoring. By contrast, the Telehealth Center of the NTUH offers a fourth-generation telehealth program that is fully integrated with remote management systems and synchronous data transfers that allow for constant analytical support backed

by full therapeutic authority and decision making from physicians outside office hours [13]. The TASMINH4 trial provides a sharp contrast to our work and confirms our belief that synchronous telehealth care that preserves people-to-people communication is the key for successful telehealth interventions. Our study results support the use of synchronous telehealth services to control home BP in patients with CVD.

Baseline BP Variability Predicts Clinical Outcomes

In our study, cumulative incidence for time to first hospitalization differed across baseline SBP-SD quartiles, implying that baseline BP variability may have some prognostic value. The Ohasama study [21] calculated day-by-day BP variability in 2455 residents in Ohasama, Japan, who measured home BP once every morning for 26 days. Over a median follow-up period of 11.9 years, an increase in SBP variability of +1 between-subject SD was found to be associated with increased hazard ratios for cardiovascular death (1.27; *P*=.002) and stroke mortality (1.41; *P*=.001). The Japan Morning Surge - Home Blood Pressure (J-HOP) study [19] calculated day-by-day BP variability in 4231 participants who measured home BP each morning and evening over a 14-day period. After 4 years of follow-up, greater day-by-day home SBP variability was associated with increased risk of incident CVD and cardiovascular end-organ damage. Both the Ohasama study and the J-HOP study concluded that baseline or short periods of day-by-day home BP variability provide useful clinical information for assessing future cardiovascular risk. In terms of physiological conditions, BP variations occur on a beat-by-beat basis and may represent a complex homeostatic control or response to neural, humoral, vascular, environmental, behavioral, and emotional stimuli [31]. Day-by-day home BP variability is not constant and can be modified: our study demonstrated that synchronous telehealth interventions can significantly decrease day-by-day BP variability. To the best of our knowledge, this is the first study that demonstrates the value of synchronous telehealth services in improving day-by-day BP variability. Future studies are warranted to determine whether declines in day-by-day BP variability could prevent future CVD occurrence.

Different Scales of BP Variability

Previous longitudinal and observational studies have shown that different types of BP variability determined by the duration of analysis can represent differing prognostic relevance for CVD and renal outcomes [22]. Durations range from very short term (beat-to-beat), short term (over 24 hours), mid-term (day-to-day), long term (visit-to-visit < 5 years), and very long term (visit-to-visit ≥ 5 years). Analysis over all durations provide a consistent message showing that higher BP variability predicts worse clinical outcomes. Higher mid-term BP variability was associated with higher rates of subclinical organ damage, stroke, myocardial infarction, cardiovascular mortality, all-cause mortality, microabuminuria, and lower estimated glomerular filtration rate [22]. The mid-term BP variability analyzed in our study showed that the cumulative incidence for time to first hospitalization is higher in patients with higher baseline BP variability, although the cumulative incidence for time to first hospitalization did not differ across subgroups stratified by hypertension status. This result may be attributed to the effects of synchronous telehealth interventions that improved BP control over time, particularly for the poorly controlled hypertension group.

The causative mechanism between BP variability and clinical outcomes is still unclear. Kario et al [32] proposed a systemic hemodynamic atherothrombotic syndrome theory that hypothesized that BP variability is caused by synergistic resonance, and demonstrated that BP exhibits different variabilities with different time phases (beat-by-beat, diurnal, day-by-day, visit-to-visit, seasonal, and yearly). Synergistic resonance of BP variabilities over different time phases can generate hemodynamic surges and trigger a vicious cycle of hemodynamic stress, which can lead to organ damage and ultimately advance to CVD.

Limitations

There were several limitations in our study. First, the study was a non-RCT, retrospective study with no comparable usual-care group. Second, although our follow-up period continued for 22 months, most patients with CVD did not receive parallel continuous telehealth services. Third, the study included heterogenous patients with CVD both with and without hypertension. To clarify specific issues, these patients were further stratified into subgroups, thus lowering the patient numbers in each subgroup. Fourth, adjustment of medication data was lacking in our study.

Conclusion

Synchronous telehealth interventions may improve home BP control and decrease day-by-day home BP variability in patients with CVD.

Acknowledgments

We thank Ms Yu-Hsien Chen for assistance in statistical analysis and the Department of Medical Research, National Taiwan University Hospital, for provision of data from the Integrated Medical Database (NTUH-iMD). This study was funded by the Ministry of Science and Technology in Taiwan (grant no. 108-2314-B-002-200).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Table S1. Change in mean blood pressure and pulse rate during the period of telehealth in subgroups.

[DOCX File , 18 KB - jmir_v24i1e22957_app1.docx]

Multimedia Appendix 2
Table S2. Change in blood pressure during the period of telehealth in all patients.
[DOCX File , 17 KB - jmir_v24i1e22957_app2.docx]

References

1. Zhou B, Bentham J, Di Cesare M, Bixby H, Danaei G, Cowan M, et al. . [doi: 10.1016/S0140-6736(16)31919-5] [Medline: 27863813]
2. Global Burden of Metabolic Risk Factors for Chronic Diseases Collaboration. Cardiovascular disease, chronic kidney disease, and diabetes mortality burden of cardiometabolic risk factors from 1980 to 2010: a comparative risk assessment. Lancet Diabetes Endocrinol 2014 Aug 01;2(8):634-647 [FREE Full text] [doi: 10.1016/S2213-8587(14)70102-0] [Medline: 24842598]
3. Lewington S, Clarke R, Qizilbash N, Peto R, Collins R, Prospective Studies Collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. Lancet 2002 Dec 14;360(9349):1903-1913. [doi: 10.1016/s0140-6736(02)11911-8] [Medline: 12493255]
4. Czernichow S, Zanchetti A, Turnbull F, Barzi F, Ninomiya T, Kengne A, Blood Pressure Lowering Treatment Trialists' Collaboration. The effects of blood pressure reduction and of different blood pressure-lowering regimens on major cardiovascular events according to baseline blood pressure: meta-analysis of randomized trials. J Hypertens 2011 Jan;29(1):4-16. [doi: 10.1097/HJH.0b013e32834000be] [Medline: 20881867]
5. Chobanian AV, Bakris GL, Black HR, Cushman WC, Green LA, Izzo JL, National Heart, Lung, Blood Institute Joint National Committee on Prevention, Detection, Evaluation,Treatment of High Blood Pressure, National High Blood Pressure Education Program Coordinating Committee. The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. JAMA 2003 May 21;289(19):2560-2572. [doi: 10.1001/jama.289.19.2560] [Medline: 12748199]
6. Whelton PK, Carey RM, Aronow WS, Casey DE, Collins KJ, Dennison Himmelfarb C, et al. 2017 ACC/AHA/AAPA/ABC/ACPM/AGS/APhA/ASH/ASPC/NMA/PCNA Guideline for the guideline for the prevention, detection, evaluation, and management of high blood pressure in adults: a report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice GuidelinesAmerican College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. J Am Coll Cardiol 2018 May 15;71(19):e127-e248 [FREE Full text] [doi: 10.1016/j.jacc.2017.11.006] [Medline: 29146535]
7. Williams B, Mancia G, Spiering W, Agabiti Rosei E, Azizi M, Burnier M, ESC Scientific Document Group. 2018 ESC/ESH Guidelines for the management of arterial hypertension. Eur Heart J 2018 Sep 01;39(33):3021-3104. [doi: 10.1093/eurheartj/ehy339] [Medline: 30165516]
8. Preis SR, Pencina MJ, Hwang S, D'Agostino RB, Savage PJ, Levy D, et al. Trends in cardiovascular disease risk factors in individuals with and without diabetes mellitus in the Framingham Heart Study. Circulation 2009 Jul 21;120(3):212-220 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.108.846519] [Medline: 19581493]
9. Yoon SSS, Carroll MD, Fryar CD. Hypertension prevalence and control among adults: United States, 2011-2014. NCHS Data Brief 2015 Nov(220):1-8 [FREE Full text] [Medline: 26633197]
10. Gee ME, Bienek A, Campbell NRC, Bancej CM, Robitaille C, Kaczorowski J, et al. Prevalence of, and barriers to, preventive lifestyle behaviors in hypertension (from a national survey of Canadians with hypertension). Am J Cardiol 2012 Feb 15;109(4):570-575. [doi: 10.1016/j.amjcard.2011.09.051] [Medline: 22154320]
11. Omboni S, Gazzola T, Carabelli G, Parati G. Clinical usefulness and cost effectiveness of home blood pressure telemonitoring: meta-analysis of randomized controlled studies. J Hypertens 2013 Mar;31(3):455-67; discussion 467. [doi: 10.1097/HJH.0b013e32835ca8dd] [Medline: 23299557]
12. Burke LE, Ma J, Azar KMJ, Bennett GG, Peterson ED, Zheng Y, et al. Current science on consumer use of mobile health for cardiovascular disease prevention. Circulation 2015 Sep 22;132(12):1157-1213. [doi: 10.1161/CIR.0000000000000232] [Medline: 26271892]
13. Anker SD, Koehler F, Abraham WT. Telemedicine and remote management of patients with heart failure. Lancet 2011 Aug 20;378(9792):731-739. [doi: 10.1016/S0140-6736(11)61229-4] [Medline: 21856487]
14. Chen YH, Lin YH, Hung CS, Huang CC, Yeih DF, Chuang PY, et al. Clinical outcome and cost-effectiveness of a synchronous telehealth service for seniors and nonseniors with cardiovascular diseases: quasi-experimental study. J Med Internet Res 2013 Apr 24;15(4):e87 [FREE Full text] [doi: 10.2196/jmir.2091] [Medline: 23615318]
15. Ho YL, Yu JY, Lin YH, Chen YH, Huang CC, Hsu TP, et al. Assessment of the cost-effectiveness and clinical outcomes of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease. J Med Internet Res 2014 Jun 10;16(6):e145 [FREE Full text] [doi: 10.2196/jmir.3346] [Medline: 24915187]

16. Hung CS, Yu JY, Lin YH, Chen YH, Huang CC, Lee JK, et al. Mortality benefit of a fourth-generation synchronous telehealth program for the management of chronic cardiovascular disease: a longitudinal study. J Med Internet Res 2016 May 13;18(5):e102 [FREE Full text] [doi: 10.2196/jmir.5718] [Medline: 27177497]
17. Hung CS, Lee JK, Chen YH, Huang CC, Wu VC, Wu HW, et al. Effect of contract compliance rate to a fourth-generation telehealth program on the risk of hospitalization in patients with chronic kidney disease: retrospective cohort study. J Med Internet Res 2018 Jan 24;20(1):e23 [FREE Full text] [doi: 10.2196/jmir.8914] [Medline: 29367185]
18. Ho TW, Huang CW, Lin CM, Lai F, Ding JJ, Ho YL, et al. A telesurveillance system with automatic electrocardiogram interpretation based on support vector machine and rule-based processing. JMIR Med Inform 2015 May 07;3(2):e21 [FREE Full text] [doi: 10.2196/medinform.4397] [Medline: 25953306]
19. Hoshide S, Yano Y, Mizuno H, Kanegae H, Kario K. Day-by-day variability of home blood pressure and incident cardiovascular disease in clinical practice: the J-HOP Study (Japan Morning Surge-Home Blood Pressure). Hypertension 2018 Jan;71(1):177-184. [doi: 10.1161/HYPERTENSIONAHA.117.10385] [Medline: 29133364]
20. Rapsomaniki E, Timmis A, George J, Pujades-Rodriguez M, Shah AD, Denaxas S, et al. Blood pressure and incidence of twelve cardiovascular diseases: lifetime risks, healthy life-years lost, and age-specific associations in 1·25 million people. Lancet 2014 May 31;383(9932):1899-1911 [FREE Full text] [doi: 10.1016/S0140-6736(14)60685-1] [Medline: 24881994]
21. Kikuya M, Ohkubo T, Metoki H, Asayama K, Hara A, Obara T, et al. Day-by-day variability of blood pressure and heart rate at home as a novel predictor of prognosis: the Ohasama study. Hypertension 2008 Dec;52(6):1045-1050. [doi: 10.1161/HYPERTENSIONAHA.107.104620] [Medline: 18981332]
22. Parati G, Ochoa JE, Bilo G, Agarwal R, Covic A, Dekker FW, European RenalCardiovascular Medicine (EURECA-m) working group of the European Renal Association-European Dialysis Transplantation Association (ERA-EDTA). Hypertension in chronic kidney disease part 2: role of ambulatory and home blood pressure monitoring for assessing alterations in blood pressure variability and blood pressure profiles. Hypertension 2016 Jun;67(6):1102-1110. [doi: 10.1161/HYPERTENSIONAHA.115.06896] [Medline: 27141057]
23. Ather S, Chan W, Chillar A, Aguilar D, Pritchett AM, Ramasubbu K, et al. Association of systolic blood pressure with mortality in patients with heart failure with reduced ejection fraction: a complex relationship. Am Heart J 2011 Mar;161(3):567-573 [FREE Full text] [doi: 10.1016/j.ahj.2010.12.009] [Medline: 21392613]
24. SPRINT Research Group, Wright JT, Williamson JD, Whelton PK, Snyder JK, Sink KM, et al. A randomized trial of intensive versus standard blood-pressure control. N Engl J Med 2015 Nov 26;373(22):2103-2116 [FREE Full text] [doi: 10.1056/NEJMoa1511939] [Medline: 26551272]
25. ACCORD Study Group, Cushman WC, Evans GW, Byington RP, Goff DC, Grimm RH, et al. Effects of intensive blood-pressure control in type 2 diabetes mellitus. N Engl J Med 2010 Apr 29;362(17):1575-1585 [FREE Full text] [doi: 10.1056/NEJMoa1001286] [Medline: 20228401]
26. Sim JJ, Zhou H, Bhandari S, Wei R, Brettler JW, Tran-Nguyen J, et al. Low systolic blood pressure from treatment and association with serious falls/syncope. Am J Prev Med 2018 Oct;55(4):488-496. [doi: 10.1016/j.amepre.2018.05.026] [Medline: 30166081]
27. Barochiner J, Alfie J, Marín MJ, Aparicio LS, Rada MA, Morales MS, et al. Prevalence and related factors of office and home hypotension in older treated hypertensive patients. Aging Clin Exp Res 2019 Jul;31(7):1011-1017. [doi: 10.1007/s40520-018-1045-3] [Medline: 30276633]
28. Divisón-Garrote JA, Ruilope LM, de la Sierra A, de la Cruz JJ, Vinyoles E, Gorostidi M, et al. Magnitude of hypotension based on office and ambulatory blood pressure monitoring: results from a cohort of 5066 treated hypertensive patients aged 80 years and older. J Am Med Dir Assoc 2017 May 01;18(5):452.e1-452.e6. [doi: 10.1016/j.jamda.2017.01.015] [Medline: 28246017]
29. Saito H, Ozaki A, Kobayashi Y, Sawano T, Tanimoto T. Pharmaceutical company payments to executive board members of professional medical associations in Japan. JAMA Intern Med 2019 Apr 01;179(4):578-580 [FREE Full text] [doi: 10.1001/jamainternmed.2018.7283] [Medline: 30715087]
30. McManus RJ, Mant J, Franssen M, Nickless A, Schwartz C, Hodgkinson J, et al. Efficacy of self-monitored blood pressure, with or without telemonitoring, for titration of antihypertensive medication (TASMINH4): an unmasked randomised controlled trial. Lancet 2018 Mar 10;391(10124):949-959 [FREE Full text] [doi: 10.1016/S0140-6736(18)30309-X] [Medline: 29499873]
31. Parati G, Ochoa JE, Lombardi C, Bilo G. Assessment and management of blood-pressure variability. Nat Rev Cardiol 2013 Mar;10(3):143-155. [doi: 10.1038/nrcardio.2013.1] [Medline: 23399972]
32. Kario K. Systemic hemodynamic atherothrombotic syndrome and resonance hypothesis of blood pressure variability: triggering cardiovascular events. Korean Circ J 2016 Jul;46(4):456-467 [FREE Full text] [doi: 10.4070/kcj.2016.46.4.456] [Medline: 27482253]

Abbreviations

ACCORD: Action to Control Cardiovascular Risk in Diabetes
ARV: average real variability

BP: blood pressure
CV: coefficient variability
CVD: cardiovascular disease
DBP: diastolic blood pressure
ESC: European Society of Cardiology
HbA_{1c}: hemoglobin A_{1c}
J-HOP: Japan Morning Surge - Home Blood Pressure
MBP: mean blood pressure
NCD: noncommunicable disease
NTUH: National Taiwan University Hospital
PR: pulse rate
RCT: randomized controlled trial
SBP: systolic blood pressure
SD: standard deviation
SPRINT: Systolic Blood Pressure Intervention Trial
TELEHEALTH: Taiwan ELEctroHEALTH

Edited by G Eysenbach; submitted 28.07.20; peer-reviewed by C Liao, M Singh; comments to author 19.08.20; revised version received 20.10.20; accepted 05.11.20; published 10.01.22.

Please cite as:
Chen YH, Hung CS, Huang CC, Lee JK, Yu JY, Ho YL
The Impact of Synchronous Telehealth Services With a Digital Platform on Day-by-Day Home Blood Pressure Variability in Patients with Cardiovascular Diseases: Retrospective Cohort Study
J Med Internet Res 2022;24(1):e22957
URL: https://www.jmir.org/2022/1/e22957
doi:10.2196/22957
PMID:35006089

Original Paper

Health Care Providers' and Professionals' Experiences With Telehealth Oncology Implementation During the COVID-19 Pandemic: A Qualitative Study

Kea Turner[1], MA, MPH, PhD; Margarita Bobonis Babilonia[2], MD; Cristina Naso[3], MEd; Oliver Nguyen[4], BS; Brian D Gonzalez[1], PhD; Laura B Oswald[1], PhD; Edmondo Robinson[5], MD, MS, MBA; Jennifer Elston Lafata[6], PhD; Robert J Ferguson[7], PhD; Amir Alishahi Tabriz[1], MD, MPH, PhD; Krupal B Patel[8], MD, MSc; Julie Hallanger-Johnson[8], MD; Nasrin Aldawoodi[9], MD; Young-Rock Hong[10], MPH, PhD; Heather S L Jim[1], PhD; Philippe E Spiess[11], MD, MS

[1]Department of Health Outcomes and Behavior, Moffitt Cancer Center, Tampa, FL, United States

[2]Department of Supportive Care Medicine, Moffitt Cancer Center, Tampa, FL, United States

[3]Virtual Health Program, Moffitt Cancer Center, Tampa, FL, United States

[4]Department of Health Outcomes and Biomedical Information, University of Florida, Gainesville, FL, United States

[5]Department of Internal and Hospital Medicine, Moffitt Cancer Center, Tampa, FL, United States

[6]Division of Pharmaceutical Outcomes and Policy, University of North Carolina at Chapel Hill, UNC Lineberger Comprehensive Cancer Center, Chapel Hill, NC, United States

[7]Department of Medicine, University of Pittsburgh, Pittsburgh, PA, United States

[8]Department of Head and Neck-Endocrine Oncology, Moffitt Cancer Center, Tampa, FL, United States

[9]Department of Anesthesiology, Moffitt Cancer Center, Tampa, FL, United States

[10]Department of Health Services Research, Management and Policy, University of Florida, Tampa, FL, United States

[11]Department of Genitourinary Oncology, Moffitt Cancer Center, Tampa, FL, United States

Corresponding Author:
Kea Turner, MA, MPH, PhD
Department of Health Outcomes and Behavior
Moffitt Cancer Center
12902 USF Magnolia Drive, MRC-CANCONT
Tampa, FL, 33612-9416
United States
Phone: 1 (813) 745 5213
Email: kea.turner@moffitt.org

Abstract

Background: Rapid implementation of telehealth for cancer care during COVID-19 required innovative and adaptive solutions among oncology health care providers and professionals (HPPs).

Objective: The aim of this qualitative study was to explore oncology HPPs' experiences with telehealth implementation during the COVID-19 pandemic.

Methods: This study was conducted at Moffitt Cancer Center (Moffitt), an NCI (National Cancer Institute)-Designated Comprehensive Cancer Center. Prior to COVID-19, Moffitt piloted telehealth visits on a limited basis. After COVID-19, Moffitt rapidly expanded telehealth visits. Telehealth visits included real-time videoconferencing between HPPs and patients and virtual check-ins (ie, brief communication with an HPP by telephone only). We conducted semistructured interviews with 40 oncology HPPs who implemented telehealth during COVID-19. The interviews were recorded, transcribed verbatim, and analyzed for themes using Dedoose software (version 4.12).

Results: Approximately half of the 40 participants were physicians (n=22, 55%), and one-quarter of the participants were advanced practice providers (n=10, 25%). Other participants included social workers (n=3, 8%), psychologists (n=2, 5%), dieticians (n=2, 5%), and a pharmacist (n=1, 3%). Five key themes were identified: (1) establishing and maintaining patient-HPP relationships, (2) coordinating care with other HPPs and informal caregivers, (3) adapting in-person assessments for telehealth, (4) developing workflows and allocating resources, and (5) future recommendations. Participants described innovative strategies for implementing

telehealth, such as coordinating interdisciplinary visits with multiple HPPs and inviting informal caregivers (eg, spouse) to participate in telehealth visits. Health care workers discussed key challenges, such as workflow integration, lack of physical exam and biometric data, and overcoming the digital divide (eg, telehealth accessibility among patients with communication-related disabilities). Participants recommended policy advocacy to support telehealth (eg, medical licensure policies) and monitoring how telehealth affects patient outcomes and health care delivery.

Conclusions: To support telehealth growth, implementation strategies are needed to ensure that HPPs and patients have the tools necessary to effectively engage in telehealth. At the same time, cancer care organizations will need to engage in advocacy to ensure that policies are supportive of oncology telehealth and develop systems to monitor the impact of telehealth on patient outcomes, health care quality, costs, and equity.

(J Med Internet Res 2022;24(1):e29635) doi:10.2196/29635

KEYWORDS

telehealth; telemedicine; teleoncology; digital health; remote monitoring; cancer; oncology; coronavirus disease; COVID-19

Introduction

The onset of the COVID-19 pandemic in March 2020 accelerated the rapid adoption of telehealth, which is health care delivery at a distance [1-4]. Prior to the COVID-19 pandemic, health care systems faced substantial barriers to telehealth adoption, such as limited reimbursement and start-up hurdles (eg, infrastructure costs) [5-9]. In the United States, telehealth had spread in some health care markets (eg, neurology) due to policy reform, while diffusion into other markets, such as oncology, was minimal [10-13]. Widespread federal and state changes to telehealth regulation, payment, and insurance benefit design as a result of COVID-19 reduced many entry barriers, allowing health care systems to deploy telehealth for oncology [4,14-16]. This seismic shift in telehealth policy created a window of opportunity to redesign cancer care for a virtual setting. In response, cancer care systems invested in large-scale organizational change, such as adopting new technologies and changing billing processes, to support telehealth implementation [17-21]. Much of this work was led by cancer care frontline workers who quickly adapted their practice to a virtual environment. The 1-year anniversary of COVID-19 was an opportunity to evaluate oncology care teams' experiences with telehealth and assess the role that telehealth can play in cancer care in the future.

Before the pandemic, the Centers for Medicare & Medicaid Services (CMS) defined telehealth as real-time videoconferencing between health care providers and professionals (HPPs) and patients [22]. During the pandemic, CMS broadened its definition of telehealth to include virtual check-ins (ie, brief communication with a practitioner by telephone or another form of telecommunication) and e-visits (ie, asynchronous communication between patients and HPPs through a secure platform, such as a patient portal). In response, oncology care teams began to implement telehealth in various forms [17-21]. In the span of just a few weeks after the pandemic hit, many cancer care systems implemented telehealth for the first time and rapidly scaled virtual care to more than half of their patient population [18-20]. HPPs were required to adapt elements of the physical exam to a telehealth visit and maintain strong relationships with patients from a distance [17,20,21,23-25]. HPPs were also faced with a long-standing digital divide, that is, disparities in patients' technology know-how and access [26]. Early reports of telehealth implementation described challenges, such as limited device access, lack of home broadband, limited digital literacy, and lack of technology education programs for patients [26]. Further, HPPs were required to make complex decisions around patient care, such as determining which conditions and patients are appropriate for virtual care [17,24].

To capture experiences with telehealth implementation during COVID-19, we conducted a qualitative study with HPPs at a cancer center. We discuss insights from oncology HPPs regarding the successes and challenges with telehealth implementation and recommendations for future telehealth delivery.

Methods

Setting, Recruitment, and Eligibility Criteria

This study was conducted at Moffitt Cancer Center (Moffitt), an NCI (National Cancer Institute)-Designated Comprehensive Cancer Center, located in Tampa, Florida. Prior to COVID-19, Moffitt piloted telehealth visits (ie, real-time videoconferencing between HPPs and patients) [22] on a limited basis. After the start of the COVID-19 pandemic, Moffitt rapidly expanded telehealth visits; the volume increased by 5000% from pre–COVID-19 to after the start of the COVID-19 pandemic [27]. Moffitt also instituted virtual check-ins (ie, brief communication with an HPP by telephone only) [22] as a part of telehealth delivery. Moffitt did not employ any other telecommunication technologies for virtual check-ins (eg, secure text messages) and did not implement any other forms of telehealth (eg, e-visits). The Moffitt Virtual Health Program—a centralized support system for the institution—provided information technology (IT) support for patients and HPPs, technical assistance for HPPs (eg, help with out-of-state licensure questions), and an interpreter service for patients with limited English proficiency. Telehealth visits were implemented through Zoom software (Zoom Video Communications, Inc), which was not integrated into Moffitt's electronic health record (EHR) system. The telehealth visits were conducted in various locations depending on what space was available to the HPP (eg, private office, shared office, and conference room). Some HPPs also elected to deliver telehealth visits from their home.

The target population included all full-time HPPs in oncology at Moffitt. HPPs were defined as health care providers who are licensed to diagnose and deliver treatment (eg, physicians, advanced practice providers [APPs], and psychologists) and professionals who are licensed to deliver other health services (eg, dieticians, social workers, and pharmacists). Oncology was defined, using the NCI's definition, as a field of medicine concerned with the diagnosis and treatment of cancer [28]. We excluded trainees (eg, fellows, residents, and interns), given that their experience with adapting to telehealth would likely be distinct from HPPs. Since trainees are new to delivering care, we opted for interviewing HPPs who had experience with delivering care prior to the pandemic and may be better positioned to describe how health care had changed because of the pandemic. For recruitment, we sent recruitment emails to two institution-wide listservs on July 1, 2021; presented the study at a leadership meeting that included all department chairs on July 15, 2021; and asked participating HPPs to suggest additional participants. We recruited until no new themes emerged or when data saturation was achieved [29].

Data Collection

The research team developed a semistructured interview guide to assess oncology HPPs' experiences with telehealth during COVID-19 (Multimedia Appendix 1). The interviews assessed how HPPs were using telehealth, adaptations necessary to shift in-person care to virtual, and perceived successes, challenges, and lessons learned. The interview guide was developed based on the Consolidated Framework for Implementation Research (CFIR) that identifies common, multilevel barriers and facilitators to implementing new programs, such as leadership support, resources, and individuals' knowledge about the intervention [30]. The guide was structured around the five CFIR domains and focused on constructs most relevant to telehealth implementation, including the following: (1) intervention characteristics (eg, ease of implementation of telehealth platform), (2) inner setting characteristics (eg, access to information about the intervention), (3) outer setting characteristics (eg, reimbursement and licensure policies related to telehealth), (4) individual characteristics (eg, patient and HPP knowledge about telehealth), and (5) process (eg, planning for telehealth implementation).

Participants were also asked about the future of oncology telehealth and whether they were supportive of using telehealth beyond COVID-19. An individual trained in qualitative methods conducted 40 individual interviews from August to December 2020 with HPPs via Zoom videoconference. An audio recording was generated through QuickTime Player (Apple Inc). We limited the interviews to 30 minutes to ensure that HPPs could participate. To stay close to the 30-minute time frame, we alerted all participants when 25 minutes had passed, assessed how many questions were left at the 25-minute mark, and if more than one question was left, we asked participants if they could stay on for a few additional minutes to finish all the interview questions. We were able to ask all participants all the interview questions using this approach. Participants provided informed consent via videoconference. Participants were asked to provide oral consent for interview participation and then asked to provide oral consent to having the interview recorded. The interviews were recorded and transcribed verbatim by a professional transcription service (GMR Transcription Services, Inc). The transcripts were deidentified (ie, all identifying information about participants was removed) and were assigned a participant ID. The Advarra Institutional Review Board reviewed and exempted this study. Participants were not compensated for their participation.

Data Analyses

We applied a hybrid approach of integrating deductive and inductive coding, which is commonly used in implementation science research and allows researchers to use theory-driven and data-driven coding [31,32]. We developed a codebook based on deductive codes [33] from the interview guide and inductive codes [34] generated from themes within the data. The interviews were coded by two independent coders using Dedoose software (version 4.12). The coders applied codes to an initial set of transcripts (n=5) independently, compared and refined coding until consensus was achieved, and finished coding the remaining transcripts. A high level of interrater reliability (κ=0.81) was achieved. The κ coefficient was defined as the number of characters coded and not coded by both coders (ie, agreement) divided by the total number of characters [35]. A summary of key themes was sent to 5 participants to confirm that the researchers' interpretation of the data was consistent with participants' experiences (ie, member checking) [29]. We reported on themes that were discussed by a minimum of 10 participants to ensure that we were summarizing central themes. Within themes, we noted when there was consensus or divergence across HPPs. The findings are reported based on the Consolidated Criteria for Reporting Qualitative Studies checklist [36].

Results

Participant Characteristics

Approximately half of the 40 participants were physicians (n=22, 55%), and one-quarter of the participants were APPs (n=10, 25%; Table 1). Other participants included social workers (n=3, 8%), psychologists (n=2, 5%), dieticians (n=2, 5%), and a pharmacist (n=1, 3%). Most participants were female (n=24, 60%). Participants represented 17 clinical areas. Few participants (n=6, 15%) had used telehealth prior to COVID-19, but all were using telehealth during COVID-19. Participants used telehealth across the cancer care continuum, including screening, diagnosis and follow-up, surveillance, supportive care, procedure preparation and follow-up, and survivorship care.

Table 1. Participant characteristics.

Characteristics	Value (N=40)
Job role, n (%)[a]	
Advanced practice provider[b]	10 (25)
Dietician	2 (5)
Pharmacist	1 (3)
Physician[c]	22 (55)
Psychologist	2 (5)
Social worker	3 (8)
Gender, n (%)	
Female	24 (60)
Male	16 (40)
Clinical focus, n (%)	
Breast oncology	2 (5)
Bone marrow transplant	2 (5)
Cutaneous oncology	2 (5)
Endocrinology	1 (3)
Gastrointestinal oncology	4 (10)
Genitourinary oncology	3 (8)
Gynecologic oncology	2 (5)
Head and neck cancer	2 (5)
Interventional radiology	2 (5)
Neuro-oncology	1 (3)
Radiation oncology	2 (5)
Sarcoma	3 (8)
Senior adult	1 (3)
Social work	3 (8)
Supportive care	4 (10)
Survivorship clinic	2 (5)
Thoracic oncology	4 (10)
Clinical affiliation, n (%)	
Single-site practice	37 (93)
Multisite practice	3 (8)
Job tenure (years), mean (SD)	12.5 (6.9)
Virtual visits pre–COVID-19, n (%)	
Yes	6 (15)
No	34 (85)
Virtual visits during COVID-19, n (%)	
Yes	40 (100)
No	0 (0)

[a]Some of the percentages may not add to 100 due to rounding.

[b]Advanced practice providers included nurse practitioners and physician assistants.

[c]Physicians included endocrinologists, medical oncologists, palliative care specialists, psychiatrists, radiation oncologists, radiologists, and surgeons.

Qualitative Analysis

Five key themes were identified: (1) establishing and maintaining patient-HPP relationships, (2) coordinating care with other HPPs and informal caregivers, (3) adapting in-person assessments for telehealth, (4) developing workflows and allocating resources, and (5) future recommendations. For each theme, the codes, code definitions, and frequency of use across participants are presented in tables in the following sections, while illustrative quotations are presented in the sections following the tables.

Theme 1: Establishing and Maintaining Patient-HPP Relationships

Overview

Participants described how telehealth changed the patient-HPP relationship, including information exchange, patient engagement, and emotional response (Table 2).

Table 2. Theme 1: establishing and maintaining patient-HPP relationships codebook.

Parent code and child codes	Code definition	Code frequency across participants (N=40), n (%)
Patient-HPP[a] communication		
Patient receptivity to information	Apply code when participant discusses how patient-HPP communication is easier due to patient's increased receptivity to receive information in their home environment.	21 (53)
Easier to share screen with patient to display results	Apply code when participant discusses how it is easier to share screen and display results with a patient during telehealth visits.	12 (30)
Patient-initiated discussion and questions	Apply code when participant describes how telehealth visits affect patients' willingness to initiate discussion about condition or ask questions.	11 (28)
Patient-HPP engagement		
Requires more energy from the HPP	Apply code when participant describes having to put on an act, be more dynamic, or put more energy into telehealth visits to engage patients.	27 (68)
Value of video	Apply code when participant describes how the video component of telehealth visits is important for patient-HPP engagement.	10 (25)
Lack of physical connection	Apply code when participant describes how lack of physical connection (ie, ability to touch patient) affects the delivery of medicine through telehealth.	12 (30)
Communicating difficult news	Apply code when participant describes the challenges of delivering difficult news through telehealth (eg, new and serious diagnosis).	23 (58)

[a]HPP: health care provider and professional.

Information Exchange

Participants felt that patient education was easier to deliver virtually because patients were more relaxed at home compared to the clinic and may be more receptive to receiving information. For example, one dietician explained as follows:

> *Because they're typically at home, they're more relaxed. They're not in this clinic environment. So, it's almost like they're a little bit more receptive to what you're saying because they're home in their own comfortable environment.*

Patient Engagement

In contrast, some participants felt that patients were more reluctant to speak up and ask questions during telehealth visits compared to in-person visits, making information exchange more difficult. One APP described this as follows:

> *I think the connection, it's not always there. It makes it a little harder to have that quick back-and-forth dialogue. When you're trying to get a lot of information across, I feel like patients are listening more than they are having a conversation back with you.*

Participants noted that patient engagement was more challenging during telehealth visits. Participants described how telehealth visits required more energy from HPPs. One APP said the following:

> *When I'm talking to patients over Zoom, those are really shorter visits. There's not as much discussion over Zoom. I fear that we're gonna miss some things because of it. I try to slow down a little bit during the Zoom visit and really try to hold their attention a little bit longer. It requires more focus on my part than an in-person visit.*

Participants also noted that engagement was harder during virtual check-ins delivered by phone compared to telehealth visits with video. One social worker indicated the following:

> *Some people they're not using their picture or they're calling in, so you don't know what's going on. So, something does get lost.*

Emotional Response

Some participants also found it more challenging to respond to patients' emotions during telehealth visits. One physician described the following:

> *I think sometimes it's nice when you can hold somebody's hand or give them a hug. Obviously, now, the COVID world, you know, we're not able to do that as much. But especially through a screen when you're not there in person, you lose a bit of that connection.*

As a result, some HPPs felt that certain tasks, such as delivering difficult news, are better suited for in-person visits. For example, one APP shared the following:

> *This is a new diagnosis. They should be in the clinic so that I can emotionally support them, offer other services while they're in here—if they need to talk to the social worker. Otherwise, they may hang up, and it may be too devastating, and we may not be able to connect again.*

Theme 2: Coordinating Care With Other HPPs and Informal Caregivers

Overview

Participants indicated that telehealth made it easier to coordinate care with other HPPs and informal caregivers, such as family members or friends (Table 3). This theme was primarily discussed by health care providers (eg, physicians and APPs) as opposed to other health care professionals (eg, dieticians).

Table 3. Theme 2: care coordination with other HPPs and informal caregivers codebook.

Parent code and child codes	Code definition	Code frequency across participants (N=40), n (%)
HPP[a]-HPP coordination		
Coordinating with external HPPs	Apply code when participant describes coordinating telehealth visits with health care HPPs external to Moffitt[b].	18 (45)
Coordinating with internal HPPs	Apply code when participant describes coordinating telehealth visits with health care HPPs internal to Moffitt (eg, other specialties).	22 (55)
HPP-caregiver coordination		
Allowing caregivers to join in-person visits through telehealth	Apply code when participant describes using Zoom or other platform to allow caregiver to participant in a patient's in-person visit.	30 (75)
Allowing caregivers to join patients' telehealth visits	Apply code when participant describes including caregivers in patient's telehealth visit.	25 (63)

[a]HPP: health care provider and professional.

[b]Moffitt: Moffitt Cancer Center.

Care Coordination With HPPs

Participants discussed how telehealth made it possible to bring together HPPs within the same institution. One physician recalled the following:

> *I've had one of the surgeons, myself, and the medical oncologist, and the radiation oncologist, and the patient all on Zoom at one time, so true multidisciplinary care provided through the Zoom platform. I've done that in one instance where the patient was in the room with me and I was doing the physical exam and we Zoomed-in the medical oncologist as part of a kind of multidisciplinary assessment.*

Participants also described how telehealth visits made it much easier to coordinate care with HPPs outside of Moffitt. For example, one physician described the following:

> *I've got a few patients that I share with oncologists in other parts of the country and there's one particularly memorable patient in my mind that I saw while he was at home as part of his visit with one of the medical oncologists at [name of organization]. So, the three of us had a three-way conversation. So, we set up interdisciplinary, interinstitutional care for that patient because he's a snowbird and winters down here and summers up there and so we shared his care. That went remarkably well.*

Care Coordination With Informal Caregivers

Participants described how telehealth made it easier to coordinate information with multiple informal caregivers (eg, inviting family members to a visit to go over a patient's prognosis). One physician said the following:

> *We used Zoom so the whole family could be there. I was able to share my screen, demonstrating the tumor. They thought that was just the greatest. And, so everybody was able to be together for all that information.*

Participants also used telehealth to enhance in-person visits, such as calling or videoconferencing caregivers who could not participate in the in-person visit due to on-site guest restrictions during COVID-19. A physician recalled the following:

> *It allowed them to have their family with them when we went over the results. Because, right now, we're still not allowing family in the outpatient clinic. And, so it allowed our patients to be with their families so that they can ask questions and hear everything at the same time.*

Theme 3: Adapting In-Person Assessments for Telehealth

Overview

Participants identified challenges with patient assessments (eg, obtaining vital signs, patient-reported outcomes [PROs], and images) and conducting physical exams during telehealth visits (Table 4).

Table 4. Theme 3: adapting in-person assessments for telehealth codebook.

Parent code and child codes	Code definition	Code frequency across participants (N=40), n (%)
Lack of physical exam		
Inability to examine lymph nodes	Apply code when participant describes the inability to feel, measure, or inspect a patient's lymph nodes during a telehealth visit.	12 (30)
Missing a clinical problem	Apply code when participant describes concerns over missing a clinical problem because of the inability to visualize the patient during a telehealth visit.	26 (65)
Getting the patient involved in the exam	Apply code when participants describe strategies for getting the patient to help with the physical exam during the telehealth visit.	16 (40)
Lack of data		
Images	Apply code when participants describe challenges with image resolution during telehealth visits to visualize condition (eg, wound monitoring).	15 (38)
Vital signs and other biometrics	Apply code when participant describes not having access to vital sign or other biometric data (eg, blood pressure) that is relevant for clinical decision-making.	22 (55)
Patient-reported outcomes	Apply code when participant describes not having access to patient-reported outcomes as a barrier for telehealth visit delivery.	11 (28)

Vital Sign Collection

Some participants lacked data due to scheduling challenges (eg, having a virtual visit scheduled before imaging data were available) or lack of remote monitoring (ie, ability to gather data between in-person visits). A few participants described how some patients would obtain their own vital sign data (eg, from wearable devices) and report it during the visit. As an example, one APP said the following:

> *It would be helpful to have vitals. I have one patient who does this by themselves, checking their own vitals (heart rate, oxygen saturation) and their blood pressure and they share it during the visit, which is great.*

Patient-Reported Outcomes

Some participants described how it was challenging to deliver certain types of care (eg, supportive care) virtually without access to PROs (eg, depression symptoms). One psychologist explained as follows:

> *We used to collect the ESAS [Edmonton Symptom Assessment System] [37] for patients in person, but now, if we see a patient virtually, we don't have that information. For supportive care, monitoring depressive symptoms is really important and now we don't have that. We talked about having nurses call the patient to collect it over the phone before the virtual visit, but that is really not the same as getting the symptoms from the patient perspective.*

During COVID-19, some clinics suspended in-person collection of PROs and did not have a means for collecting PROs as a part of virtual care. Participants also described how image quality could be challenging, requiring HPPs to follow up and obtain images after the visit. One physician indicated the following:

> *Some patients might have a limitation in the internet provider bandwidth, so the image resolution is poor quality. In that circumstance, what I do is I tell the patient, "Please take a picture of your surgical site with your phone, and email it to me." But it would be nice if there was a more systematic way to do this so I could see the image during the visit.*

Image Collection

Participants felt that most patients were comfortable sharing images with HPPs digitally and did not cite concerns about privacy. A few HPPs noted that patients over 65 years of age were more reluctant to share images due to privacy concerns compared to younger patients. One APP said the following:

> *Our clinic sees a lot of older patients [over 65 years] who tend to have more privacy issues with sharing images than our younger patients.*

Physical Exams

Participants were also concerned about missing important clinical problems due to the lack of a physical exam (eg, checking lymph nodes). One APP said the following:

> *I really do think there's something lost, the personal touch and the things you see with your eyes and a physical exam for a cancer patient is very important. To feel for lymph adenopathy and wherever they say their cancer was, the nearest lymph node drainage.*

> *I mean, there's no way that I could discern that over telemedicine.*

Similarly, a physician explained as follows:

> *I see things on my patients all the time when I see them in person. I'm like, "You need to go get that little thing on your arm there checked." It might be a new skin lesion that they need to have looked at. I worry about missing things during virtual visits when I can't really visualize the patient.*

Some participants described getting the patient to help with conducting the physical exam (eg, having the patient measure a visible tumor) during the telehealth visit to overcome this limitation. For example, one APP shared the following:

> *Sometimes, I've had patients who've had visible tumors on their neck or visible tumors, I can see it on telemedicine, but I can't really measure it. So I ask the patient to go get a tape measure and measure it.*

Theme 4: Developing Workflows and Allocating Resources

Overview

Participants described the importance of developing workflows for telehealth visits that are equivalent to in-person visits (eg, check-in process, EHR integration, and scheduling) and ensure sufficient resources are allocated for HPPs and patients (Table 5).

Table 5. Theme 4: developing workflows and allocating resources for telehealth codebook.

Parent code and child codes	Code definition	Code frequency across participants (N=40), n (%)
Workflow		
Check-in process	Apply code when participant describes the check-in process used during telehealth visits.	28 (70)
Scheduling	Apply code when participants describe how telehealth visits are scheduled (eg, batching visits).	33 (83)
Electronic health record (EHR) integration	Apply code when participant describes lack of Zoom EHR integration (eg, inability to find visit in EHR).	16 (40)
Resources		
Equipment	Apply code when participants describe equipment necessary for telehealth delivery (eg, cameras).	35 (88)
Space	Apply code when participants describe space where telehealth visit is conducted.	36 (90)
Clerical support	Apply code when participants describe the amount of administrative support available for telehealth visit delivery.	26 (65)
Information technology (IT) support for patients with low digital literacy	Apply code when participants discuss IT support for patients who may have low computer or mobile health literacy.	27 (68)
Tools for patients with disabilities	Apply code when participants discuss need for tools for patients with disabilities (eg, closed captioning and speech-to-text tools).	30 (75)

Check-in Process

Participants, for example, felt that the patient check-in process, which is usually handled by staff, was left up to the HPP for telehealth visits, creating inefficiency. One physician described the following:

> *I really am concerned about the way we've implemented the virtual visits is we've cut the nurse and the PAR [Patient Access Representative] team out of the equation so now I'm having to check in my own patients when they show up on a Zoom visit. That becomes a little more cumbersome because any time you ask me to do an administrative task, the chance that I'm gonna execute it accurately is less than if you have a nurse doing it.*

EHR Integration

Participants also found it challenging that telehealth visits were not integrated within the EHR, making it difficult to ensure that all the HPPs involved in preparing a visit (eg, nurse and APP) could access the visit link. One APP mentioned the following:

> *I have great nurses who work with me and they prep the clinics and they don't have the Zoom invitation. So, the invitation comes to me, but it doesn't go to the nurses working with me, which makes it hard for them to prep the visit. But if somehow, the Zoom meetings were discoverable in the EHR then that would really help.*

Scheduling

Some participants described refining their schedule to ensure that telehealth visits were grouped together rather than interspersed between in-person visits. Having telehealth and in-person visits mixed together created problems, such as forcing HPPs to run back and forth between their clinic and office where telehealth visits were conducted. One physician mentioned the following:

> *So, for a month or two I was running around like a chicken without a head. And finally, I said, "Enough. We're gonna have designated days where we're doing all the virtual visits." So, I can sit here and do one patient after the other because it was not working, running back and forth between clinics and my office where I take the virtual visits.*

HPP Resources

Some participants described lacking sufficient resources for telehealth visits, including equipment (eg, cameras and headphones), space (eg, finding a private or well-lit space), and administrative support. One social worker shared the following:

> *How are we supposed to do virtual visits when we don't have a camera? Here we are, five or six months into this pandemic and most of us have no access to a camera outside of our cellphones.*

Participants also felt that the amount of administrative support for telehealth visits was inferior to in-person visits and, as a result, much of the administrative work fell on APPs and nurses. One APP mentioned the following:

> *Right now, our nurse is looking ahead at who's scheduled and making sure everything is there for the visit. But, it's almost like if there could be someone on the back end—like they do for new patient visits that are in person—doing that, it would help the nurses. It ends up being a lot of clerical work for the nurses, taking them away from patient education.*

Patient Resources

Most participants felt that there were sufficient resources for patients, such as the interpreter service and IT support. Participants felt that the level of IT support worked for the majority of patients but was not sufficient for patients with low mobile health (mHealth) literacy or those with a lower ability to use mHealth apps with efficiency to accomplish a task [38]. As a result, some HPPs spent a significant amount of time helping their patients with the Zoom app during telehealth visits or converted telehealth visits to virtual check-ins by phone. One pharmacist described the following:

> *The patients that struggled the most were patients who had to access the telehealth visit from their phone. They didn't know how to download or locate the app. We need to make sure the training that patients receive covers how to use the app so that we don't have to spend so much time on this during the visit.*

Participants also indicated a need for more resources for patients with disabilities that may interfere with technology use. One psychologist described the following:

> *It was really challenging to meet virtually with patients who had trouble with hearing. I ended up converting those visits [to in person].*

Similarly, a physician described the difficulty of delivering telehealth visits to patients with speech impairments, as follows:

> *We see patients who may not be able to talk after surgery. We would try to get the caregiver on or use the chat feature, but if they don't have help at home, we'd have to bring them in.*

Theme 5: Recommendations for Telehealth Implementation in the Future

Patient-Level Recommendations

Nearly all participants were supportive of continuing telehealth beyond COVID-19 but recommended changes to ensure implementation is sustainable (Table 6). Participants provided recommendations at the patient, HPP, and organizational levels for improving telehealth implementation in the future. At the patient level, participants discussed the importance of overcoming the digital divide and recommended real-time IT support for patients, closed captioning for patients with communication-related disabilities, and educational materials on how to prepare for a telehealth visit (eg, finding a place that is comfortable). For example, one psychologist suggested the following:

> *I would have said a bit more patient training, just some training on not just using Zoom, but telemedicine in general, and sort of optimizing the whole visit because I'm sure there are ways to do it. For example, encouraging them to find a comfortable space, have what they need on hand.*

Table 6. Recommendations for future telehealth implementation codebook.

Parent code and child codes	Code definition	Code frequency across participants (N=40), n (%)
Patient-level recommendations		
Greater telehealth accessibility	Apply code when participants recommend strategies for improving the accessibility of telehealth (eg, closed captioning).	31 (78)
Real-time, information technology support	Apply code when participants recommend strategies to deliver more timely technology support to patients.	12 (30)
Patient education	Apply code when participants recommend strategies to deliver more patient education on how to use telehealth or prepare for telehealth visits.	23 (58)
HPP[a]-level recommendations		
Sharing information about telehealth policy changes	Apply code when participants recommend strategies to promote discussion about ongoing telehealth policy changes (eg, licensure).	21 (53)
Sharing best practices	Apply code when participants recommend strategies to promote discussion about telehealth best practices (eg, tips for patient engagement and checklists).	33 (83)
"Webside manner" training	Apply code when participants recommend HPP-level training to engage patients in a telehealth environment (eg, how to maintain eye contact).	32 (80)
Production support	Apply code when participants recommend support needed to professionalize the telehealth visit (eg, background and lighting).	26 (65)
Organizational-level recommendations		
Optimizing workflow	Apply code when participants recommend strategies for optimizing workflow (eg, virtual waiting room and batching telehealth visits).	30 (75)
Policy advocacy	Apply code when participants recommend organizational strategies for policy advocacy (eg, being more engaged with advocacy organizations).	10 (25)
Long-term planning	Apply code when participants recommend long-term planning strategies, such as how telehealth will be evaluated and how it will fit with other organizational priorities.	20 (50)

[a]HPP: health care provider and professional.

HPP-Level Recommendations

Participants thought that HPPs should have more training on engaging patients in a virtual environment (eg, optimal eye contact and communication strategies to create a dialogue). One APP mentioned the following:

> *I think more training would be helpful. Not on the technology, but what is the best way to maintain eye contact with patients, how do you keep patients engaged and talking with you.*

Participants also wanted more discussion about best practices in telehealth use and updates on policy changes (eg, Medicare reimbursement). One physician said the following:

> *I wish there was more ongoing communication, not just deliver the news without an ongoing conversation. When Medicare is going to stop covering all the telemedicine visits is a question that we all have heard. It would be helpful to have a chance to discuss this.*

Participants also wanted more resources to ensure that telehealth visits were delivered in a professional and consistent way across HPPs. For example, one physician shared the following:

> *Professionalizing it is the highest priority. Making sure the equipment is in place, that a dedicated room is in place, that the background is place, the lighting is in place. I think we should be trying to go in a direction more like a professional broadcasting company with that level of quality because I think that adds to the performance art that is at the heart of a lot of medicine.*

Organizational-Level Recommendations

At the organizational level, participants suggested that Moffitt engage in telehealth policy advocacy (eg, reimbursement and out-of-state licensure). One physician recalled the following:

> *What we had asked that patient to do was to drive themselves across the border and to have the telemedicine visit from their car. I mean, it's an absurdity. There's nothing we can actionably do about*

> *that except work on our government relations team and kind of change that into the future.*

Participants also had suggestions for optimizing workflow, such as having a virtual waiting room and letting patients know when an HPP is running late for a telehealth visit. For example, one APP shared the following:

> *When I go to my cardiologist, they have a virtual waiting room that is managed by their MAs [medical assistants]. So, I log in and the MA immediately says to me, "Hi. Glad you're with us. The doctor is running 10 minutes behind." In our scenario, if I'm not on time, the patients think there's something wrong with the technology and they'll be calling the nurse. They'll be calling like crazy. And it's just another thing for the nurse to have to pay attention to.*

Participants also recommended developing a long-term strategy for telehealth, such as how telehealth aligns with other organizational priorities and an evaluation plan to see how telehealth affects patient outcomes and health care quality and costs. One physician shared the following:

> *I'd like to see a long-term vision and plans for evaluating progress. We don't really know how it will impact patients or the care we deliver or reimbursement. We should see if it is being used to its full potential in other priority areas, like clinical trials.*

Discussion

Principal Findings

The goal of this study was to capture oncology HPPs' experiences with rapid implementation of telehealth during the COVID-19 pandemic. Overall, oncology HPPs saw telehealth as an integral part of health care delivery moving forward—a finding consistent with other studies [20,24]—but recommended addressing key barriers to improve sustainability. At Moffitt, we plan to present the findings of this research to our leadership team to determine how the results can inform future telehealth implementation. More broadly, our research findings also provide implications for future telehealth research and practice. For example, our findings suggest that more work is needed to overcome the digital divide and ensure that HPPs have access to the resources and data necessary to deliver high-quality care in a virtual environment (eg, vital signs). Further, the results suggest that a long-term strategy is needed to determine how telehealth will be integrated across the cancer care continuum and monitored to assess impact on patient outcomes and health care delivery. At the same time, health care systems will need to develop a research and policy agenda to ensure that evidence informs telehealth policy approaches, and that the regulatory and payment landscape accelerates and facilitates optimal telehealth use in cancer care.

Addressing the Digital Divide

Health care systems have implemented innovative strategies to address the digital divide (ie, disparities in technology access, skills, and use) during COVID-19, such as assessing patients' readiness for telehealth [39-42]. A key step to overcoming the digital divide is understanding which patients within a system are impacted by the digital divide and what specific barriers they experience. For example, smartphone-only internet access may bridge the digital divide for some patients, but it could limit health care access for others who have limited data plans or limited experience with mobile apps [18,43]. To address digital health literacy (ie, the ability to use computers and search for and evaluate health information electronically), some health care systems have deployed social work staff to help patients access telehealth [18,41]. In our system, we provided IT support to all patients with scheduled telehealth visits, but for some patients, more support was needed. Health care systems have also experimented with device loan programs and partnering with community-based organizations to create spaces where patients can connect to the internet [21,41,43]. The Veterans Health Administration (VHA) recently developed a partnership with a cellphone carrier to ensure patients could access VHA telehealth apps without affecting their data limits [21]. Prior studies also suggest that experience with technology is a key predictor of technology use for health care (eg, how often a patient uses the internet or a smartphone) [44]. Further studies should explore longitudinal models of technology training that move beyond a one-time training and allow for repeated experiences with technology use for health care. Further, our health care team members found that current telehealth apps may not be optimized for patients with disabilities (eg, lack of closed captioning and speech-to-text tools), similar to other studies [21,45,46]. To improve access among patients with disabilities and other conditions that could affect human-computer interaction (eg, low literacy), technology vendors should include patients who are affected by the digital divide in co-designing and testing of new products [39,46,47]. Future studies should also consider community-based participatory research approaches in the development of digital health technologies to better engage patients affected by the digital divide [48]. Community-based participatory research has been used in mHealth studies, for example, to increase community participation in app development, usability testing and app refinement, and designing recruitment, implementation, and dissemination strategies [48].

Our study participants noted that access to interpreters was a key ingredient of successful telehealth deployment. Prior studies have documented disparities in telehealth use among patients who prefer English and those who do not prefer English in the United States during the pandemic [49-52]. Similar disparities have been observed in patient portals, which are used by some health care systems to deliver telehealth [53,54]. Qualitative studies have documented barriers, such as limited access to professional interpreters, lack of bilingual HPPs, and lack of translation of COVID-19–related informational materials [55]. Additional studies are needed to better understand telehealth access among patients who do not prefer English. For example, a previous study demonstrated that factors such as interpreter modality (eg, professional vs ad hoc and video vs in person) affect the accuracy of interpretation for health care visits [56]. Future studies could compare different models of interpreter services through telehealth and compare interpretation accuracy rates and other outcomes, such as patient satisfaction. Researchers have also recommended that health care

organizations develop monitoring systems for evaluating disparities in telehealth uptake by language preference (eg, dashboards), develop telehealth and patient portal trainings in multiple languages, and prioritize the hiring of bilingual HPPs [57]. More research is needed to develop and test implementation strategies that address disparities in telehealth access based on language preference.

Ensuring HPPs Have Sufficient Tools for Telehealth Implementation

HPPs may need additional resources to deliver telehealth effectively. HPP-facing tools [58,59] may be helpful in providing education on "webside manner" and implementation checklists [58,59]. HPPs at Moffitt indicated a need for guidance on the optimal way to conduct a telehealth visit (eg, lighting and eye contact), similar to other studies [40,60]. Health care systems could disseminate available trainings (eg, Academy of Communication in Healthcare training) or develop institution-specific trainings [61]. Some health care systems have developed implementation checklists for health care team members that include helpful tips, such as confirming a patient's phone number at the beginning of the virtual visit in case the technology fails [62]. Patient-facing tools may also improve virtual patient-HPP communication. Some HPPs have developed patient handouts on how to conduct elements of a physical exam during a telehealth visit and how to prepare for the visit (eg, patient positioning) [23]. Similar approaches could be tested more broadly. Other studies have noted that HPPs may lack sufficient digital health literacy [63,64], which can negatively affect engagement with health IT systems (eg, EHRs) [65]. This concern was not raised by our participants, but future studies should explore the effects of HPPs' eHealth literacy on telehealth implementation.

Remote Monitoring for Telehealth Implementation

Access to patient data during telehealth visits, such as biometric data and PROs, is also critical to implementation success. A recent study in primary care found that blood pressure assessments declined by 37% during COVID-19, in part due to lack of biometric screening during virtual visits [66,67]. These findings highlight the need for remote monitoring programs, which will require health care systems to invest in complex change (eg, EHR integration of biometric data and optimized data visualization) [68-72]. Prior to COVID-19, there was limited reimbursement for remote monitoring, hindering health care system adoption [73]. During the pandemic, Medicare has expanded payment policies for remote monitoring with certain restrictions (eg, type of data and minimum amount of data needed) [74]. Some cancer care systems have started to invest in remote monitoring programs and cited recent changes in Medicare policy as a motivator for adoption [75]. Cancer care systems will need to evaluate the effectiveness of remote monitoring programs and identify areas in which remote monitoring adds the most value. Like other studies, our research found that older patients may have more privacy concerns about remote monitoring and sharing patient-generated health data, such as images, compared to other patients [76,77]. Prior studies recommended strategies that strengthen patient activation and HPP-patient trust; they also recommended developing patient education programs about how data are being used and protected in order to overcome privacy-related barriers to sharing patient-generated data [76,77]. Other privacy-related barriers (eg, concerns about information security) were not mentioned by participants in our study but deserve consideration in future telehealth research. Organizations have reported examples of "Zoombombing," or when an intruder joins a Zoom videoconference [78], raising concerns about information security and telehealth. Researchers have recommended that health care organizations develop multipronged approaches (eg, employee training and simulated cyberattacks) to address information security threats in telehealth [79]. Further, some have argued that health care organizations should transition from consumer-oriented videoconference tools that were adopted at the onset of the pandemic to health care–specific videoconference tools with additional security features [79]. More research is needed to identify best practices in information security for telehealth as it grows in usage.

Greater Evidence Regarding Effectiveness, Implementation, and Potential Risks of Telehealth

Cancer care systems, payers and insurers, and policy makers will need more evidence for the effectiveness and safety of oncology telehealth to guide future decisions. Research from other health care sectors has demonstrated that telehealth can be equivalent to in-person care for certain conditions and offers a relative advantage over in-person care (eg, reducing rural health care disparities) [11,80-83]. Telehealth also has potential risks, such as inappropriate antibiotic prescribing or exacerbating existing health care disparities due to the digital divide [26,84-89]. Within oncology, telehealth models for supportive and survivorship care and ancillary services (eg, genetic counseling) have proven effective, but there is limited evaluation of telehealth for other areas of care (eg, screening, diagnosis, treatment, and surveillance) [90-95]. Therefore, it is critical to evaluate telehealth use in these areas and assess impact on patient outcomes, health care quality, costs, equity, and potential risks (eg, inappropriate care). Further, research is needed to evaluate strategies for incentivizing the use of telehealth in a postpandemic landscape [96]. One strategy may be alternative payment models, which have increased telehealth adoption in other health care sectors but are understudied in oncology [97]. Further, as adoption decisions move from mandatory to voluntary, studies should test theories of technology adoption (eg, technology acceptance model) to examine what factors help explain sustained telehealth use beyond the pandemic [98,99].

Policy Advocacy for Telehealth

Cancer care systems and key stakeholders will need to develop an agenda to ensure that future policies are supportive of oncology telehealth. During the pandemic, many state-level restrictions were lifted that made it easier to implement telehealth, including waiving out-of-state licensure requirements or expanding payment parity [4,15]. Further, many state Medicaid programs and commercial insurers changed telehealth policies in response to COVID-19 (eg, removing cost-sharing requirements) [4,15]. At the federal level, there was a major overhaul of Medicare telehealth payment policies (eg, allowing virtual check-ins through telephone to qualify for telehealth),

resulting in 244 temporary regulatory changes [4,15,100]. Uncertainty remains regarding which federal- and state-level policies will be retained in the future [4,15]. There has also been federal investment in overcoming the digital divide in the United States [101]. The Infrastructure Investment and Jobs Act became public law recently and provides funding for expanding broadband access in low-income neighborhoods, reducing practices of digital redlining (ie, limited internet service provision in low-income and high–minority concentration neighborhoods), and expanding internet subsidies for individuals with limited economic resources [101]. Future research will be needed to monitor program implementation and effectiveness for addressing digital disparities at the federal, state, and local levels. Moving forward, a policy agenda will need to include greater investment in telehealth research, addressing medical licensure and credentialing barriers [102], and a balanced approach to regulation, one that continues to fuel innovation in telehealth for oncology while safeguarding against potential risks (eg, delivering telehealth for a condition that is not "tele-amenable" [96]). For instance, among our study participants, some HPPs expressed concerns about using telehealth to deliver information about a new and serious cancer diagnosis. A recent survey among oncologists (n=29) identified similar findings: some oncologists were reluctant to use telehealth for delivering bad news [103]. Additional studies are needed to determine optimal use of telehealth in oncology to guide future policy.

Limitations

This paper has several limitations. First, this is a qualitative study from an NCI-Designated Comprehensive Cancer Center and the findings may not be generalizable to other settings. Second, there are other health care professionals (eg, genetic counselors) who have used telehealth as a part of their practice, and their experiences were not captured in our sample. Third, we limited the interviews to 30 minutes or less to increase participation, and we reduced the number of questions included in the interview guide. Therefore, some topics, such as how HPPs have used telehealth for clinical trials, were not discussed. Fourth, our study excluded residents and fellows who were important stakeholders in telehealth implementation. Future studies should examine the unique experiences of residents and fellows who are simultaneously learning how to deliver care in person and virtually. Finally, it was beyond the scope of this research to capture the patient perspective. Studies have started to assess patient experience with telehealth during the COVID-19 pandemic; however, there has been limited study of this in oncology [104-106]. Future studies should assess the perspectives of patients with cancer regarding telehealth to explore patient satisfaction, barriers and facilitators to telehealth access, and patient preferences for telehealth (eg, whether certain services should be delivered in person vs virtually and preferences for how interpreter services should be implemented).

Conclusions

Overall, cancer care frontline HPPs have used innovative and adaptive strategies to rapidly implement telehealth during COVID-19 and are supportive of continuing virtual cancer care delivery beyond the COVID-19 pandemic. HPPs identified several facilitators for telehealth implementation, such as improved care coordination with other HPPs and informal caregivers. HPPs also noted several barriers, such as lack of physical examinations and vital sign information, which limited HPPs' ability to fully evaluate a patient. To support the rapid growth of oncology telehealth, implementation strategies are needed to overcome the digital divide and ensure that HPPs and patients have the tools necessary to effectively engage in telehealth. Health care systems, policy makers, health insurers, and payers must develop long-term strategies for integrating telehealth into the cancer care continuum, building the evidence base around telehealth in oncology, and developing a policy agenda that will advance telehealth innovation while safeguarding against potential risks.

Authors' Contributions

KT developed the research question and study design, assisted with the qualitative analyses, and drafted the manuscript. MBB, CN, and ON helped refine the interview guide, assisted with conducting qualitative interviews, assisted with conducting the qualitative analyses, reviewed the manuscript draft, and provided feedback. BDG, LBO, ER, JEL, RJF, AAT, KBP, JHJ, NA, YRH, and HSLJ reviewed the summary report of the qualitative data analyses and the manuscript draft and provided feedback. PES helped refine aspects of the study design and methodology, provided input on the interview guide, assisted with participant recruitment, reviewed the manuscript draft, and provided feedback.

Conflicts of Interest

BDG has received funding from SureMed Compliance and KenPharm. JHJ is a consultant for HRA Pharma for unrelated work. HSLJ has received funding from RedHill BioPharma, Janssen Scientific Affairs, and Merck.

Multimedia Appendix 1
Interview guide.
[DOCX File , 19 KB - jmir_v24i1e29635_app1.docx]

References

1. Fagherazzi G, Goetzinger C, Rashid MA, Aguayo GA, Huiart L. Digital health strategies to fight COVID-19 worldwide: Challenges, recommendations, and a call for papers. J Med Internet Res 2020 Jun 16;22(6):e19284 [FREE Full text] [doi: 10.2196/19284] [Medline: 32501804]
2. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Trends in outpatient care delivery and telemedicine during the COVID-19 pandemic in the US. JAMA Intern Med 2021 Mar 01;181(3):388-391. [doi: 10.1001/jamainternmed.2020.5928] [Medline: 33196765]
3. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Variation in telemedicine use and outpatient care during the COVID-19 pandemic in the United States. Health Aff (Millwood) 2021 Feb;40(2):349-358. [doi: 10.1377/hlthaff.2020.01786] [Medline: 33523745]
4. Mehrotra A, Bhatia RS, Snoswell CL. Paying for telemedicine after the pandemic. JAMA 2021 Feb 02;325(5):431-432. [doi: 10.1001/jama.2020.25706] [Medline: 33528545]
5. LeRouge C, Garfield MJ. Crossing the telemedicine chasm: Have the US barriers to widespread adoption of telemedicine been significantly reduced? Int J Environ Res Public Health 2013 Nov 28;10(12):6472-6484 [FREE Full text] [doi: 10.3390/ijerph10126472] [Medline: 24287864]
6. Rogove HJ, McArthur D, Demaerschalk BM, Vespa PM. Barriers to telemedicine: Survey of current users in acute care units. Telemed J E Health 2012;18(1):48-53. [doi: 10.1089/tmj.2011.0071] [Medline: 22082107]
7. Weinstein RS, Lopez AM, Joseph BA, Erps KA, Holcomb M, Barker GP, et al. Telemedicine, telehealth, and mobile health applications that work: Opportunities and barriers. Am J Med 2014 Mar;127(3):183-187. [doi: 10.1016/j.amjmed.2013.09.032] [Medline: 24384059]
8. Scott KC, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: A systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
9. Moffatt JJ, Eley DS. Barriers to the up-take of telemedicine in Australia--A view from providers. Rural Remote Health 2011;11(2):1581 [FREE Full text] [Medline: 21385004]
10. Richard JV, Wilcock AD, Schwamm LH, Uscher-Pines L, Zachrison KS, Siddiqui A, et al. Assessment of telestroke capacity in US hospitals. JAMA Neurol 2020 Aug 01;77(8):1035-1037. [doi: 10.1001/jamaneurol.2020.1274] [Medline: 32453424]
11. Zhang D, Wang G, Zhu W, Thapa JR, Switzer JA, Hess DC, et al. Expansion of telestroke services improves quality of care provided in super rural areas. Health Aff (Millwood) 2018 Dec;37(12):2005-2013. [doi: 10.1377/hlthaff.2018.05089] [Medline: 30633675]
12. Pareek P, Vishnoi JR, Kombathula SH, Vyas RK, Misra S. Teleoncology: The youngest pillar of oncology. JCO Glob Oncol 2020 Sep;6:1455-1460 [FREE Full text] [doi: 10.1200/GO.20.00295] [Medline: 32997540]
13. Shea CM, Turner K, Alishahi Tabriz A, North S. Implementation strategies for telestroke: A qualitative study of telestroke networks in North Carolina. Telemed J E Health 2019 Aug;25(8):708-716 [FREE Full text] [doi: 10.1089/tmj.2018.0131] [Medline: 30192206]
14. Weigel G, Ramaswamy A, Sobel L, Salganicoff A, Cubanski J, Freed M. Opportunities and barriers for telemedicine in the US during the COVID-19 emergency and beyond. KFF. Washington, DC: Kaiser Family Foundation; 2020 May 11. URL: https://www.kff.org/womens-health-policy/issue-brief/opportunities-and-barriers-for-telemedicine-in-the-u-s-during-the-covid-19-emergency-and-beyond/ [accessed 2022-01-10]
15. Royce TJ, Sanoff HK, Rewari A. Telemedicine for cancer care in the time of COVID-19. JAMA Oncol 2020 Nov 01;6(11):1698-1699. [doi: 10.1001/jamaoncol.2020.2684] [Medline: 32672821]
16. Kircher SM, Mulcahy M, Kalyan A, Weldon CB, Trosman JR, Benson AB. Telemedicine in oncology and reimbursement policy during COVID-19 and beyond. J Natl Compr Canc Netw 2020 Sep 30:1-7. [doi: 10.6004/jnccn.2020.7639] [Medline: 32998106]
17. Kang JJ, Wong RJ, Sherman EJ, Rybkin A, McBride SM, Riaz N, et al. The 3 Bs of cancer care amid the COVID-19 pandemic crisis: "Be safe, be smart, be kind"-A multidisciplinary approach increasing the use of radiation and embracing telemedicine for head and neck cancer. Cancer 2020 Sep 15;126(18):4092-4104 [FREE Full text] [doi: 10.1002/cncr.33031] [Medline: 32639615]
18. Knudsen KE, Willman C, Winn R. Optimizing the use of telemedicine in oncology care: Postpandemic opportunities. Clin Cancer Res 2021 Feb 15;27(4):933-936. [doi: 10.1158/1078-0432.CCR-20-3758] [Medline: 33229457]
19. Lonergan PE, Washington III SL, Branagan L, Gleason N, Pruthi RS, Carroll PR, et al. Rapid utilization of telehealth in a comprehensive cancer center as a response to COVID-19: Cross-sectional analysis. J Med Internet Res 2020 Jul 06;22(7):e19322 [FREE Full text] [doi: 10.2196/19322] [Medline: 32568721]
20. Darcourt JG, Aparicio K, Dorsey PM, Ensor JE, Zsigmond EM, Wong ST, et al. Analysis of the implementation of telehealth visits for care of patients with cancer in Houston during the COVID-19 pandemic. JCO Oncol Pract 2021 Jan;17(1):e36-e43. [doi: 10.1200/OP.20.00572] [Medline: 33026951]
21. Jiang CY, El-Kouri NT, Elliot D, Shields J, Caram MEV, Frankel TL, et al. Telehealth for cancer care in veterans: Opportunities and challenges revealed by COVID. JCO Oncol Pract 2021 Jan;17(1):22-29. [doi: 10.1200/OP.20.00520] [Medline: 32970512]

22. Medicare telemedicine health care provider fact sheet. CMS. Baltimore, MD: Centers for Medicare and Medicaid Services; 2020 Mar 17. URL: https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet [accessed 2022-01-10]
23. Prasad A, Brewster R, Newman JG, Rajasekaran K. Optimizing your telemedicine visit during the COVID-19 pandemic: Practice guidelines for patients with head and neck cancer. Head Neck 2020 Jun;42(6):1317-1321 [FREE Full text] [doi: 10.1002/hed.26197] [Medline: 32343458]
24. Chang PJ, Jay GM, Kalpakjian C, Andrews C, Smith S. Patient and provider-reported satisfaction of cancer rehabilitation telemedicine visits during the COVID-19 pandemic. PM R 2021 Dec;13(12):1362-1368 [FREE Full text] [doi: 10.1002/pmrj.12552] [Medline: 33455066]
25. Elkaddoum R, Haddad FG, Eid R, Kourie HR. Telemedicine for cancer patients during COVID-19 pandemic: Between threats and opportunities. Future Oncol 2020 Jun;16(18):1225-1227 [FREE Full text] [doi: 10.2217/fon-2020-0324] [Medline: 32356460]
26. Tam S, Wu VF, Williams AM, Girgis M, Sheqwara JZ, Siddiqui F, et al. Disparities in the uptake of telemedicine during the COVID-19 surge in a multidisciplinary head and neck cancer population by patient demographic characteristics and socioeconomic status. JAMA Otolaryngol Head Neck Surg 2021 Feb 01;147(2):209-211. [doi: 10.1001/jamaoto.2020.3052] [Medline: 33151289]
27. Drees J. Moffitt Cancer Center's virtual visits up 5,000% in response to COVID-19. Becker's Hospital Review. Chicago, IL: Becker's Hospital Review; 2020 May 15. URL: https://www.beckershospitalreview.com/oncology/moffitt-cancer-center-s-virtual-visits-up-5-000-in-response-to-covid-19.html [accessed 2022-01-10]
28. Oncology. National Cancer Institute. Bethesda, MD: National Cancer Institute URL: https://www.cancer.gov/publications/dictionaries/cancer-terms/def/oncology [accessed 2022-01-10]
29. Creswell JW, Miller DL. Determining validity in qualitative inquiry. Theory Pract 2000 Aug;39(3):124-130. [doi: 10.1207/s15430421tip3903_2]
30. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: A consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
31. Hamilton AB, Finley EP. Qualitative methods in implementation research: An introduction. Psychiatry Res 2019 Oct;280:112516 [FREE Full text] [doi: 10.1016/j.psychres.2019.112516] [Medline: 31437661]
32. Fereday J, Muir-Cochrane E. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. Int J Qual Methods 2016 Nov 29;5(1):80-92 [FREE Full text] [doi: 10.1177/160940690600500107]
33. Boyatzis RE. Transforming Qualitative Information: Thematic Analysis and Code Development. Thousand Oaks, CA: SAGE Publications, Inc; 1998.
34. Crabtree BF, Miller WL. A template approach to text analysis: Developing and using codebooks. In: Crabtree BF, Miller WL, editors. Doing Qualitative Research. 2nd edition. Thousand Oaks, CA: SAGE Publications, Inc; 1999:163-177.
35. Brennan PF, Hays BJ. The kappa statistic for establishing interrater reliability in the secondary analysis of qualitative clinical data. Res Nurs Health 1992 Apr;15(2):153-158. [doi: 10.1002/nur.4770150210] [Medline: 1565808]
36. Tong A, Sainsbury P, Craig J. Consolidated Criteria for Reporting Qualitative Research (COREQ): A 32-item checklist for interviews and focus groups. Int J Qual Health Care 2007 Dec;19(6):349-357. [doi: 10.1093/intqhc/mzm042] [Medline: 17872937]
37. Bruera E, Kuehn N, Miller MJ, Selmser P, Macmillan K. The Edmonton Symptom Assessment System (ESAS): A simple method for the assessment of palliative care patients. J Palliat Care 1991;7(2):6-9. [Medline: 1714502]
38. Lin TTC, Bautista JR. Understanding the relationships between mHealth apps' characteristics, trialability, and mHealth literacy. J Health Commun 2017 Dec;22(4):346-354. [doi: 10.1080/10810730.2017.1296508] [Medline: 28323546]
39. Rodriguez JA, Clark CR, Bates DW. Digital health equity as a necessity in the 21st Century Cures Act era. JAMA 2020 Jun 16;323(23):2381-2382. [doi: 10.1001/jama.2020.7858] [Medline: 32463421]
40. Lau J, Knudsen J, Jackson H, Wallach AB, Bouton M, Natsui S, et al. Staying connected in the COVID-19 pandemic: Telehealth at the largest safety-net system in the United States. Health Aff (Millwood) 2020 Aug;39(8):1437-1442. [doi: 10.1377/hlthaff.2020.00903] [Medline: 32525705]
41. Nouri S, Khoong E, Lyles C, Karliner L. Addressing equity in telemedicine for chronic disease management during the Covid-19 pandemic. NEJM Catal Innov Care Deliv 2020 May 04;1(3):1-13 [FREE Full text] [doi: 10.1056/CAT.20.0123]
42. Elston Lafata J, Smith AB, Wood WA, Fitzpatrick B, Royce TJ. Virtual visits in oncology: Enhancing care quality while designing for equity. JCO Oncol Pract 2021 May;17(5):220-223. [doi: 10.1200/OP.20.00645] [Medline: 33539180]
43. DeGuzman PB, Bernacchi V, Cupp CA, Dunn B, Ghamandi BJF, Hinton ID, et al. Beyond broadband: Digital inclusion as a driver of inequities in access to rural cancer care. J Cancer Surviv 2020 Oct;14(5):643-652. [doi: 10.1007/s11764-020-00874-y] [Medline: 32390103]
44. Potdar R, Thomas A, DiMeglio M, Mohiuddin K, Djibo DA, Laudanski K, et al. Access to internet, smartphone usage, and acceptability of mobile health technology among cancer patients. Support Care Cancer 2020 Nov;28(11):5455-5461. [doi: 10.1007/s00520-020-05393-1] [Medline: 32166381]

45. Annaswamy TM, Verduzco-Gutierrez M, Frieden L. Telemedicine barriers and challenges for persons with disabilities: COVID-19 and beyond. Disabil Health J 2020 Oct;13(4):100973 [FREE Full text] [doi: 10.1016/j.dhjo.2020.100973] [Medline: 32703737]
46. Valdez RS, Rogers CC, Claypool H, Trieshmann L, Frye O, Wellbeloved-Stone C, et al. Ensuring full participation of people with disabilities in an era of telehealth. J Am Med Inform Assoc 2021 Feb 15;28(2):389-392 [FREE Full text] [doi: 10.1093/jamia/ocaa297] [Medline: 33325524]
47. Crawford A, Serhal E. Digital health equity and COVID-19: The innovation curve cannot reinforce the social gradient of health. J Med Internet Res 2020 Jun 02;22(6):e19361 [FREE Full text] [doi: 10.2196/19361] [Medline: 32452816]
48. Brewer LC, Fortuna KL, Jones C, Walker R, Hayes SN, Patten CA, et al. Back to the future: Achieving health equity through health informatics and digital health. JMIR Mhealth Uhealth 2020 Jan 14;8(1):e14512 [FREE Full text] [doi: 10.2196/14512] [Medline: 31934874]
49. Rodriguez JA, Saadi A, Schwamm LH, Bates DW, Samal L. Disparities in telehealth use among California patients with limited English proficiency. Health Aff (Millwood) 2021 Mar;40(3):487-495. [doi: 10.1377/hlthaff.2020.00823] [Medline: 33646862]
50. Eberly LA, Kallan MJ, Julien HM, Haynes N, Khatana SAM, Nathan AS, et al. Patient characteristics associated with telemedicine access for primary and specialty ambulatory care during the COVID-19 pandemic. JAMA Netw Open 2020 Dec 01;3(12):e2031640 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.31640] [Medline: 33372974]
51. Lott A, Campbell KA, Hutzler L, Lajam CM. Telemedicine utilization at an academic medical center during COVID-19 pandemic: Are some patients being left behind? Telemed J E Health 2022 Jan;28(1):44-50. [doi: 10.1089/tmj.2020.0561] [Medline: 33794135]
52. Schenker RB, Laguna MC, Odisho AY, Okumura MJ, Burnett H. Are we reaching everyone? A cross-sectional study of telehealth inequity in the COVID-19 pandemic in an urban academic pediatric primary care clinic. Clin Pediatr (Phila) 2022 Jan;61(1):26-33 [FREE Full text] [doi: 10.1177/00099228211045809] [Medline: 34514898]
53. Ancker JS, Barrón Y, Rockoff ML, Hauser D, Pichardo M, Szerencsy A, et al. Use of an electronic patient portal among disadvantaged populations. J Gen Intern Med 2011 Oct;26(10):1117-1123 [FREE Full text] [doi: 10.1007/s11606-011-1749-y] [Medline: 21647748]
54. Ancker J, Nosal S, Hauser D, Way C, Calman N. Access policy and the digital divide in patient access to medical records. Health Policy Technol 2017 Mar;6(1):3-11. [doi: 10.1016/j.hlpt.2016.11.004]
55. Kucirek NK, Thomas NJ, Norman JS, Athavale P, Jaradeh K, Hsiang EY, et al. Stories from COVID-19 reveal hospitalized patients with limited English proficiency have always been uniquely prone to social isolation. J Gen Intern Med 2021 Mar;36(3):786-789 [FREE Full text] [doi: 10.1007/s11606-020-06383-z] [Medline: 33409888]
56. Nápoles A, Santoyo-Olsson J, Karliner L, Gregorich S, Pérez-Stable E. Inaccurate language interpretation and its clinical significance in the medical encounters of Spanish-speaking Latinos. Med Care 2015;53(11):940-947. [doi: 10.1097/mlr.0000000000000422]
57. Rodriguez JA, Casillas A, Cook BL, Marlin RP. The language of equity in digital health: Prioritizing the needs of limited English proficient communities in the patient portal 2.0. J Health Care Poor Underserved 2021;32(2S):211-219. [doi: 10.1353/hpu.2021.0059]
58. Chua IS, Jackson V, Kamdar M. Webside manner during the COVID-19 pandemic: Maintaining human connection during virtual visits. J Palliat Med 2020 Nov;23(11):1507-1509. [doi: 10.1089/jpm.2020.0298] [Medline: 32525744]
59. West HJ. Telemedicine in oncology: Delivering on an overdue promise in the COVID-19 era. Front Oncol 2020;10:578888 [FREE Full text] [doi: 10.3389/fonc.2020.578888] [Medline: 33102236]
60. Zulman DM, Verghese A. Virtual care, telemedicine visits, and real connection in the era of COVID-19: Unforeseen opportunity in the face of adversity. JAMA 2021 Feb 02;325(5):437-438. [doi: 10.1001/jama.2020.27304] [Medline: 33528520]
61. Telehealth communication: Quick tips. Academy of Communication in Healthcare. Lexington, KY: Academy of Communication in Healthcare URL: http://www.achonline.org/Telehealth [accessed 2022-01-10]
62. Banerjee SC, Staley JM, Howell F, Malling C, Moreno A, Kotsen C, et al. Communicating effectively via tele-oncology (Comskil TeleOnc): A guide for best practices for communication skills in virtual cancer care. J Cancer Educ 2021 Feb 05:1-6 [FREE Full text] [doi: 10.1007/s13187-021-01959-7] [Medline: 33544315]
63. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
64. Kirchberg J, Fritzmann J, Weitz J, Bork U. eHealth literacy of German physicians in the pre-COVID-19 era: Questionnaire study. JMIR Mhealth Uhealth 2020 Oct 16;8(10):e20099 [FREE Full text] [doi: 10.2196/20099] [Medline: 33064102]
65. Kuek A, Hakkennes S. Healthcare staff digital literacy levels and their attitudes towards information systems. Health Informatics J 2020 Mar;26(1):592-612 [FREE Full text] [doi: 10.1177/1460458219839613] [Medline: 30983476]
66. Alexander GC, Tajanlangit M, Heyward J, Mansour O, Qato DM, Stafford RS. Use and content of primary care office-based vs telemedicine care visits during the COVID-19 pandemic in the US. JAMA Netw Open 2020 Oct 01;3(10):e2021476 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.21476] [Medline: 33006622]

67. Herzer KR, Pronovost PJ. Ensuring quality in the era of virtual care. JAMA 2021 Feb 02;325(5):429-430. [doi: 10.1001/jama.2020.24955] [Medline: 33528544]
68. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable health technology and electronic health record integration: Scoping review and future directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
69. Zhu H, Colgan J, Reddy M, Choe EK. Sharing patient-generated data in clinical practices: An interview study. In: Proceedings of the American Medical Informatics Association Annual Symposium. 2016 Presented at: American Medical Informatics Association Annual Symposium; November 12-16, 2016; Chicago, IL p. 1303-1312 URL: http://europepmc.org/abstract/MED/28269928
70. Simblett S, Greer B, Matcham F, Curtis H, Polhemus A, Ferrão J, et al. Barriers to and facilitators of engagement with remote measurement technology for managing health: Systematic review and content analysis of findings. J Med Internet Res 2018 Jul 12;20(7):e10480 [FREE Full text] [doi: 10.2196/10480] [Medline: 30001997]
71. Rodbard D. Continuous glucose monitoring: A review of successes, challenges, and opportunities. Diabetes Technol Ther 2016 Dec;18 Suppl 2:S3-S13 [FREE Full text] [doi: 10.1089/dia.2015.0417] [Medline: 26784127]
72. Abdolkhani R, Gray K, Borda A, DeSouza R. Patient-generated health data management and quality challenges in remote patient monitoring. JAMIA Open 2019 Dec;2(4):471-478 [FREE Full text] [doi: 10.1093/jamiaopen/ooz036] [Medline: 32025644]
73. Health Care: Telehealth and Remote Patient Monitoring Use in Medicare and Selected Federal Programs. GAO-17-365. Washington, DC: Government Accountability Office; 2017 Apr. URL: https://www.gao.gov/assets/gao-17-365.pdf [accessed 2022-01-10]
74. Final policy, payment, and quality provisions changes to the Medicare physician fee schedule for calendar year 2021. CMS. 2020 Dec 01. URL: https://www.cms.gov/newsroom/fact-sheets/final-policy-payment-and-quality-provisions-changes-medicare-physician-fee-schedule-calendar-year-1 [accessed 2022-01-10]
75. Wicklund E. Mount Sinai extends cancer care to the home with new telehealth program. mHealthIntelligence. Danvers, MA: Xtelligent Healthcare Media; 2021 Jan 20. URL: https://mhealthintelligence.com/news/mount-sinai-extends-cancer-care-to-the-home-with-new-telehealth-program [accessed 2022-01-10]
76. Karampela M, Ouhbi S, Isomursu M. Connected health user willingness to share personal health data: Questionnaire study. J Med Internet Res 2019 Nov 27;21(11):e14537 [FREE Full text] [doi: 10.2196/14537] [Medline: 31774410]
77. Serrano KJ, Yu M, Riley WT, Patel V, Hughes P, Marchesini K, et al. Willingness to exchange health information via mobile devices: Findings from a population-based survey. Ann Fam Med 2016;14(1):34-40 [FREE Full text] [doi: 10.1370/afm.1888] [Medline: 26755781]
78. Ling C, Balcı U, Blackburn J, Stringhini G. A first look at Zoombombing. In: Proceedings of the IEEE Symposium on Security and Privacy. 2021 Presented at: IEEE Symposium on Security and Privacy; May 24-27, 2021; San Francisco, CA p. 1452-1467. [doi: 10.1109/sp40001.2021.00061]
79. Jalali MS, Landman A, Gordon WJ. Telemedicine, privacy, and information security in the age of COVID-19. J Am Med Inform Assoc 2021 Mar 01;28(3):671-672 [FREE Full text] [doi: 10.1093/jamia/ocaa310] [Medline: 33325533]
80. Kepplinger J, Barlinn K, Deckert S, Scheibe M, Bodechtel U, Schmitt J. Safety and efficacy of thrombolysis in telestroke: A systematic review and meta-analysis. Neurology 2016 Sep 27;87(13):1344-1351. [doi: 10.1212/WNL.0000000000003148] [Medline: 27566746]
81. Albritton J, Maddox L, Dalto J, Ridout E, Minton S. The effect of a newborn telehealth program on transfers avoided: A multiple-baseline study. Health Aff (Millwood) 2018 Dec;37(12):1990-1996. [doi: 10.1377/hlthaff.2018.05133] [Medline: 30633672]
82. Pekmezaris R, Tortez L, Williams M, Patel V, Makaryus A, Zeltser R, et al. Home telemonitoring in heart failure: A systematic review and meta-analysis. Health Aff (Millwood) 2018 Dec;37(12):1983-1989. [doi: 10.1377/hlthaff.2018.05087] [Medline: 30633680]
83. Dharmar M, Kuppermann N, Romano PS, Yang NH, Nesbitt TS, Phan J, et al. Telemedicine consultations and medication errors in rural emergency departments. Pediatrics 2013 Dec;132(6):1090-1097. [doi: 10.1542/peds.2013-1374] [Medline: 24276844]
84. Mehrotra A, Paone S, Martich GD, Albert SM, Shevchik GJ. A comparison of care at e-visits and physician office visits for sinusitis and urinary tract infection. JAMA Intern Med 2013 Jan 14;173(1):72-74 [FREE Full text] [doi: 10.1001/2013.jamainternmed.305] [Medline: 23403816]
85. Ray KN, Shi Z, Gidengil CA, Poon SJ, Uscher-Pines L, Mehrotra A. Antibiotic prescribing during pediatric direct-to-consumer telemedicine visits. Pediatrics 2019 May;143(5):e20182491 [FREE Full text] [doi: 10.1542/peds.2018-2491] [Medline: 30962253]
86. Rodriguez JA, Betancourt JR, Sequist TD, Ganguli I. Differences in the use of telephone and video telemedicine visits during the COVID-19 pandemic. Am J Manag Care 2021 Jan;27(1):21-26 [FREE Full text] [doi: 10.37765/ajmc.2021.88573] [Medline: 33471458]
87. Barnett ML, Ray KN, Souza J, Mehrotra A. Trends in telemedicine use in a large commercially insured population, 2005-2017. JAMA 2018 Nov 27;320(20):2147-2149 [FREE Full text] [doi: 10.1001/jama.2018.12354] [Medline: 30480716]

88. Mehrotra A, Jena AB, Busch AB, Souza J, Uscher-Pines L, Landon BE. Utilization of telemedicine among rural Medicare beneficiaries. JAMA 2016 May 10;315(18):2015-2016 [FREE Full text] [doi: 10.1001/jama.2016.2186] [Medline: 27163991]
89. Douglas MD, Xu J, Heggs A, Wrenn G, Mack DH, Rust G. Assessing telemedicine utilization by using Medicaid claims data. Psychiatr Serv 2017 Feb 01;68(2):173-178 [FREE Full text] [doi: 10.1176/appi.ps.201500518] [Medline: 27691381]
90. Freeman LW, White R, Ratcliff CG, Sutton S, Stewart M, Palmer JL, et al. A randomized trial comparing live and telemedicine deliveries of an imagery-based behavioral intervention for breast cancer survivors: Reducing symptoms and barriers to care. Psychooncology 2015 Aug;24(8):910-918 [FREE Full text] [doi: 10.1002/pon.3656] [Medline: 25146413]
91. Ferguson RJ, Sigmon ST, Pritchard AJ, LaBrie SL, Goetze RE, Fink CM, et al. A randomized trial of videoconference-delivered cognitive behavioral therapy for survivors of breast cancer with self-reported cognitive dysfunction. Cancer 2016 Jun 01;122(11):1782-1791 [FREE Full text] [doi: 10.1002/cncr.29891] [Medline: 27135464]
92. Pang L, Liu Z, Lin S, Liu Z, Liu H, Mai Z, et al. The effects of telemedicine on the quality of life of patients with lung cancer: A systematic review and meta-analysis. Ther Adv Chronic Dis 2020;11:1-12 [FREE Full text] [doi: 10.1177/2040622320961597] [Medline: 33101621]
93. van den Brink JL, Moorman PW, de Boer MF, Hop WCJ, Pruyn JFA, Verwoerd CDA, et al. Impact on quality of life of a telemedicine system supporting head and neck cancer patients: A controlled trial during the postoperative period at home. J Am Med Inform Assoc 2007;14(2):198-205 [FREE Full text] [doi: 10.1197/jamia.M2199] [Medline: 17213498]
94. Schwartz MD, Valdimarsdottir HB, Peshkin BN, Mandelblatt J, Nusbaum R, Huang A, et al. Randomized noninferiority trial of telephone versus in-person genetic counseling for hereditary breast and ovarian cancer. J Clin Oncol 2014 Mar 01;32(7):618-626 [FREE Full text] [doi: 10.1200/JCO.2013.51.3226] [Medline: 24449235]
95. Hilgart JS, Hayward JA, Coles B, Iredale R. Telegenetics: A systematic review of telemedicine in genetics services. Genet Med 2012 Sep;14(9):765-776. [doi: 10.1038/gim.2012.40] [Medline: 22498847]
96. Agrawal S, Ghandi T. Telehealth should be expanded—If it can address today's health care challenges. Health Affairs. Bethesda, MD: Project HOPE; 2020 Sep 23. URL: https://www.healthaffairs.org/do/10.1377/hblog20200916.264569/full/ [accessed 2022-01-10]
97. Zhao M, Hamadi H, Haley DR, Xu J, White-Williams C, Park S. Telehealth: Advances in alternative payment models. Telemed J E Health 2020 Dec;26(12):1492-1499. [doi: 10.1089/tmj.2019.0294] [Medline: 32101100]
98. Kulviwat S, Bruner II GC, Kumar A, Nasco SA, Clark T. Toward a unified theory of consumer acceptance technology. Psychol Mark 2007 Dec;24(12):1059-1084. [doi: 10.1002/mar.20196]
99. Lee Y, Kozar KA, Larsen KR. The technology acceptance model: Past, present, and future. Commun Assoc Inf Syst 2003;12:752-780. [doi: 10.17705/1cais.01250]
100. LaPointe J. Biden administration inherits slew of Medicare regulatory changes. RevCycleIntelligence. Danvers, MA: Xtelligent Healthcare Media; 2021 Jan 26. URL: https://revcycleintelligence.com/news/biden-administration-inherits-slew-of-medicare-regulatory-changes [accessed 2022-01-10]
101. H.R.3684 - Infrastructure Investment and Jobs Act. Congress.gov. 2021. URL: https://www.congress.gov/bill/117th-congress/house-bill/3684/text [accessed 2022-01-10]
102. Bell DL, Katz MH. Modernize medical licensing, and credentialing, too-Lessons from the COVID-19 pandemic. JAMA Intern Med 2021 Mar 01;181(3):312-315. [doi: 10.1001/jamainternmed.2020.8705] [Medline: 33439217]
103. Heyer A, Granberg RE, Rising KL, Binder AF, Gentsch AT, Handley NR. Medical oncology professionals' perceptions of telehealth video visits. JAMA Netw Open 2021 Jan 04;4(1):e2033967 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.33967] [Medline: 33443581]
104. Singh A, Mountjoy N, McElroy D, Mittal S, Al Hemyari B, Coffey N, et al. Patient perspectives with telehealth visits in cardiology during COVID-19: Online patient survey study. JMIR Cardio 2021 Jan 22;5(1):e25074 [FREE Full text] [doi: 10.2196/25074] [Medline: 33385234]
105. Auchus IC, Jaradeh K, Tang A, Marzan J, Boslett B. Transitioning to telehealth during the COVID-19 pandemic: Patient perspectives and attendance at an HIV clinic in San Francisco. AIDS Patient Care STDS 2021 Jul;35(7):249-254. [doi: 10.1089/apc.2021.0075] [Medline: 34242090]
106. Holcomb D, Faucher M, Bouzid J, Quint-Bouzid M, Nelson DB, Duryea E. Patient perspectives on audio-only virtual prenatal visits amidst the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic. Obstet Gynecol 2020;136(2):317-322. [doi: 10.1097/aog.0000000000004026]

Abbreviations

APP: advanced practice provider
CFIR: Consolidated Framework for Implementation Research
CMS: Centers for Medicare & Medicaid Services
EHR: electronic health record
ESAS: Edmonton Symptom Assessment System
HPP: health care provider and professional
IT: information technology

MA: medical assistant
mHealth: mobile health
Moffitt: Moffitt Cancer Center
NCI: National Cancer Institute
PAR: Patient Access Representative
PRO: patient-reported outcome
VHA: Veterans Health Administration

Edited by T Leung; submitted 14.04.21; peer-reviewed by T Haddad, T Busse; comments to author 05.05.21; revised version received 07.05.21; accepted 15.12.21; published 19.01.22.

Please cite as:
Turner K, Bobonis Babilonia M, Naso C, Nguyen O, Gonzalez BD, Oswald LB, Robinson E, Elston Lafata J, Ferguson RJ, Alishahi Tabriz A, Patel KB, Hallanger-Johnson J, Aldawoodi N, Hong YR, Jim HSL, Spiess PE
Health Care Providers' and Professionals' Experiences With Telehealth Oncology Implementation During the COVID-19 Pandemic: A Qualitative Study
J Med Internet Res 2022;24(1):e29635
URL: https://www.jmir.org/2022/1/e29635
doi:10.2196/29635
PMID:34907900

Original Paper

Telehealth Availability and Use of Related Technologies Among Medicare-Enrolled Cancer Survivors: Cross-sectional Findings From the Onset of the COVID-19 Pandemic

Yuki Lama[1*], MPH, PhD; Amy J Davidoff[2*], MS, PhD; Robin C Vanderpool[1*], DrPH; Roxanne E Jensen[3*], PhD

[1]Health Communication and Informatics Research Branch, Behavioral Research Program, National Cancer Institute, Bethesda, MD, United States
[2]Healthcare Assessment Research Branch, Healthcare Delivery Research Program, National Cancer Institute, Bethesda, MD, United States
[3]Outcomes Research Branch, Healthcare Delivery Research Program, National Cancer Institute, Bethesda, MD, United States
[*]all authors contributed equally

Corresponding Author:
Roxanne E Jensen, PhD
Outcomes Research Branch
Healthcare Delivery Research Program
National Cancer Institute
9609 Medical Center Drive
Bethesda, MD, 20892-9761
United States
Phone: 1 240 276 7588
Email: roxanne.jensen@nih.gov

Abstract

Background: There has been rapid integration of telehealth into care delivery during the COVID-19 pandemic. However, little is known about technology ownership, internet access and use for communication, and telehealth availability among cancer survivors, particularly those enrolled in Medicare.

Objective: This study aims to identify sociodemographic associations with technology ownership, internet access and use for communication, and telehealth availability in a population-based sample of Medicare-enrolled cancer survivors.

Methods: Data are from the Medicare Current Beneficiary Survey COVID-19 Summer 2020 Supplement administered between June 10 and July 15, 2020. Analyses were restricted to beneficiaries who reported a prior (nonskin) cancer diagnosis and a usual source of care (N=2044). Dichotomous outcomes included technology ownership, internet access, internet use for communication, and telehealth availability from providers. Sociodemographic correlates included sex, age, race/ethnicity, Medicare/Medicaid dual enrollment, rurality, census region, and self-reported comorbidities.

Results: Over half (957/2044, 53%) of cancer survivors reported using the internet for communication purposes, and 62% (1218/2044) reported that their usual provider had telehealth services available. Using the internet for communication purposes was reported less frequently for rural compared to urban survivors (adjusted probability of 28% vs 46%; $P<.001$) and for Hispanic and Black survivors compared to non-Hispanic White survivors (29%, 31%, and 44%, respectively; all $P<.01$). Rural survivors reported lower telehealth availability (53% vs 63%; $P<.001$); no significant differences in telehealth availability were identified by race/ethnicity.

Conclusions: During the COVID-19 pandemic, study findings highlight a complex digital divide among Medicare beneficiaries with a history of cancer related to device ownership necessary for telehealth, internet access and use for communication, and reports of providers having telehealth available. Multilevel approaches are needed to increase equitable telehealth availability and use for cancer survivors. Suggested strategies include increasing broadband internet access to providers and patients in at-risk communities, supporting telehealth implementation among providers that serve populations with known health disparities, raising awareness of providers' available telehealth services among patients, and screening for technology use and provision of telehealth-related technical assistance among older and historically underserved cancer survivors.

(J Med Internet Res 2022;24(1):e34616) doi:10.2196/34616

KEYWORDS

cancer survivor; Medicare; telehealth; COVID-19; availability; use; elderly; older adults; cancer; sociodemographic; internet; communication; population; access

Introduction

Throughout the COVID-19 pandemic, there has been a substantial increase in telehealth, defined as "the exchange of medical information from one site to another through electronic communication to improve a patient's health," to deliver health care services [1-3]. Increased telehealth use was motivated by the need to limit COVID-19 exposure, particularly among vulnerable patient populations, and facilitated by increased familiarity with teleconferencing and changes in reimbursement and regulatory policies [2]. Cancer survivors are likely to be particularly vulnerable to access to care obstacles associated with COVID-19: they tend to be older and may have comorbidities or immunosuppression that heightens both risk and consequences of infection [4]. As a result, telehealth can provide particular benefit by reducing logistical barriers to timely cancer-related care [5]. Prior research has identified barriers to successful telehealth use for adults generally, including lack of computer and internet access, and limited digital literacy skills needed to negotiate video log-on processes [6]. Previous research demonstrates that characteristics such as older age, rurality, and lower income were associated with less access to and engagement with telehealth prepandemic [7-9]. However, assessing whether these patterns apply to cancer survivors, and persisted during the onset of the pandemic, remains underexplored. Therefore, this exploratory study's aim was to identify sociodemographic correlates of technology ownership, internet access, internet use for communication, and telehealth availability, captured early during the COVID-19 pandemic among a nationally representative sample of Medicare beneficiaries with a cancer history in the United States.

Methods

Data Source and Sample

The Medicare Current Beneficiary Survey COVID-19 Summer 2020 Supplement, sponsored by the Centers for Medicare and Medicaid Services, is a telephone survey of community-dwelling Medicare beneficiaries administered between June 10 to July 15, 2020. Individuals who are 65 years or older, disabled, or have end-stage renal disease (ESRD) are eligible for Medicare [10]. Survey weights represent the population of beneficiaries continuously enrolled in Medicare from January through summer 2020. Selected beneficiaries reported prior non–skin cancer diagnosis and a usual source of care other than urgent care or emergency departments.

Outcome Variables

We evaluated four dichotomous measures: (1) technology ownership ("Do you own or use any of the following: desktop or laptop, smartphone, or tablet?"), (2) internet access ("Do you have access to the internet?"), (3) internet use for communication ("Have you ever participated in video or voice calls or conferencing over the internet, such as with Skype or FaceTime?"), and (4) telehealth availability ("Does your usual provider offer telephone or video appointments, so that you don't need to physically visit their office or facility?").

Independent Variables

We assessed the role of sociodemographic characteristics, including age, race/ethnicity, Medicare/Medicaid dual enrollment (an indicator of poverty) [11], and rurality defined by metropolitan statistical area, adjusting for sex, census region, and self-reported comorbidities.

Statistical Analyses

Descriptive statistics and bivariate comparisons were generated for each outcome. Multivariable logistic regression models estimated the effects of the independent variables on each outcome. Results related to age, race/ethnicity, Medicare/Medicaid dual enrollment, and rurality are reported as adjusted predicted (marginal) probabilities. All estimates were weighted, and analyses used SAS 9.4 (SAS Institute, Inc) procedures to adjust for complex survey design (PROC SURVEYFREQ, SURVEYMEANS, and SURVEYLOGISTIC). Statistical tests were 2-sided with α=.05. As Medicare Current Beneficiary Survey data are publicly available and deidentified, the study was not considered to be human participant research.

Results

The sample of Medicare-enrolled cancer survivors was 57% (1144/2044) female, 41% (1192/2044) 75 years or older, and 79% (1638/2044) non-Hispanic White (n=2044; weighted n=9,941,910). Over half (957/2044, 53%) used the internet for communication and 62% (1218/2044) reported telehealth availability (Table 1). Please note that the percentages reported are weighted, while the sample sizes are unweighted, and hence, percentages may differ.

Table 1. Sample characteristics of Medicare beneficiaries reporting a cancer history and a usual source of care as assessed by the Medicare Current Beneficiary Survey (N=2044).

Characteristic	Unweighted n	Weighted %
Sex		
Male	900	43.1
Female	1144	56.9
Age group (years)		
<65	201	10.8
65-74	651	48.1
≥75	1192	41.1
Race/ethnicity		
White non-Hispanic	1638	79.1
Black non-Hispanic	135	7.7
Hispanic	159	6.1
Other/unknown	112	7.1
Dual Medicare/Medicaid enrollment (2019)		
Nondual Medicare/Medicaid enrollment	1747	87.7
Any dual Medicare/Medicaid enrollment	297	12.3
Metropolitan statistical area residence[a]		
Urban	1542	79.6
Rural	502	20.4
Census region		
Northeast	372	18.2
Midwest	461	21.3
South	819	40.0
West	392	20.6
Comorbidities		
0 conditions	458	24.2
1 condition	653	32.3
≥2 conditions	933	43.5
Outcomes		
Technology ownership (computer, smartphone, tablet)	1588	83.3
Internet access	1604	82.9
Internet use for communication	957	53.0
Telehealth availability	1218	62.0

[a]Metropolitan statistical area residence defined by the Office of Management and Budget as having at least one urbanized area with a minimum population of 50,000.

Technology Ownership, Internet Access, and Internet Use for Communication

Older age, rural residence, dual Medicare/Medicaid enrollment, and non-Hispanic Black or Hispanic race/ethnicity were associated with lower probabilities of owning technology (Table 2), internet access (Table 3), and internet use for communication (all P<.05; Table 4). Compared to urban cancer survivors, rural survivors had lower predicted probabilities of technology ownership (67% vs 82%; P<.001; Table 2), internet access (58% vs 79%; P<.001; Table 3), and internet use for communication (28% vs 46%; P<.001; Table 4). Compared to non-Hispanic Whites, Hispanic and Black survivors had lower technology ownership (67% vs 82%, P<.001; 65% vs 82%, P=.005, respectively), internet access (56% vs 81%, P<.001; 52% vs 82%, P<.001, respectively), and internet use for communication (29% vs 44%, P<.001; 31% vs 44%, P=.002, respectively). Compared to nondual enrolled, dual Medicare/Medicaid enrolled

beneficiaries had lower technology ownership (60% vs 86%; *P*<.001), internet access (53% vs 83%; *P*<.001), and internet use for communication (26% vs 48%; *P*<.001).

Table 2. Factors associated with technology ownership among Medicare Current Beneficiary Survey respondents with a history of cancer and a usual source of care (N=2044).

Characteristic	Own computer, tablet, or smartphone		
	Predicted probability	aOR[a] (95% CI)	*P* value
Metropolitan statistical area residence[b]			
Urban (reference)	0.82	N/A[c]	N/A
Rural	0.67	0.44 (0.31-0.62)	<.001
Age group (years)			
65-74 (reference)	0.79	N/A	N/A
<65	0.88	2.06 (1.07-3.94)	.03
≥75	0.50	0.27 (0.19-0.39)	<.001
Sex			
Male (reference)	0.72	N/A	N/A
Female	0.75	1.14 (0.89-1.47)	.32
Race/ethnicity			
White non-Hispanic (reference)	0.82	N/A	N/A
Black non-Hispanic	0.65	0.40 (0.21-0.75)	.005
Hispanic	0.67	0.44 (0.26-0.75)	<.001
Other/unknown	0.83	1.05 (0.51-2.13)	.90
Dual Medicare/Medicaid enrollment (2019)			
Nondual Medicare/Medicaid enrollment (reference)	0.86	N/A	N/A
Any enrollment Medicare/Medicaid	0.60	0.25 (0.15-0.41)	<.001
Census region			
Midwest (reference)	0.71	N/A	N/A
Northeast	0.72	1.05 (0.68-1.63)	.83
South	0.72	1.09 (0.74-1.60)	.66
West	0.84	2.23 (1.41-3.53)	<.001
Comorbidities			
0 (reference)	0.77	N/A	N/A
1	0.74	0.85 (0.56-1.29)	.45
≥2	0.74	0.85 (0.60-1.21)	.36

[a]aOR: adjusted odds ratio.

[b]Metropolitan statistical area residence defined by the Office of Management and Budget as having at least one urbanized area with a minimum population of 50,000.

[c]N/A: not applicable.

Table 3. Factors associated with internet access among Medicare Current Beneficiary Survey respondents with a history of cancer and a usual source of care (N=2044).

Characteristic	Internet access		
	Predicted probability	aOR[a] (95% CI)	*P* value
Metropolitan statistical area residence[b]			
Metro (reference)	0.79	N/A[c]	N/A
Nonmetro	0.58	0.35 (0.25-0.50)	<.001
Age group (years)			
65-74 (reference)	0.74	N/A	N/A
<65	0.83	1.73 (0.84-3.55)	.14
≥75	0.46	0.30 (0.22-0.42)	<.001
Sex			
Male (reference)	0.70	N/A	N/A
Female	0.69	0.92 (0.73-1.17)	.49
Race/ethnicity			
White non-Hispanic (reference)	0.81	N/A	N/A
Black non-Hispanic	0.52	0.26 (0.13-0.52)	<.001
Hispanic	0.56	0.30 (0.18-0.51)	<.001
Other/unknown	0.82	1.04 (0.55-1.94)	.91
Dual Medicare/Medicaid enrollment (2019)			
Nondual Medicare/Medicaid enrollment (reference)	0.83	N/A	N/A
Any enrollment Medicare/Medicaid	0.53	0.24 (0.16-0.34)	<.001
Census region			
Midwest (reference)	0.65	N/A	N/A
Northeast	0.63	0.94 (0.60-1.45)	.76
South	0.67	1.11 (0.73-1.71)	.62
West	0.81	2.35 (1.29-4.28)	.01
Comorbidities			
0 (reference)	0.71	N/A	N/A
1	0.69	0.91 (0.61-1.34)	.62
≥2	0.68	0.88 (0.58-1.32)	.52

[a]aOR: adjusted odds ratio.

[b]Metropolitan statistical area residence defined by the Office of Management and Budget as having at least one urbanized area with a minimum population of 50,000.

[c]N/A: not applicable.

Table 4. Factors associated with internet use for communication among Medicare Current Beneficiary Survey respondents with a history of cancer and a usual source of care (N=2044).

Characteristic	Internet use for communication		
	Predicted probability	aOR[a] (95% CI)	*P* value
Metropolitan statistical area residence[b]			
Metro (reference)	0.46	N/A[c]	N/A
Nonmetro	N/A	0.45 (0.34-0.59)	<.001
Age group (years)			
65-74 (reference)	0.42	N/A	N/A
<65	0.47	1.22 (0.84-1.77)	.30
≥75	0.22	0.38 (0.31-0.47)	<.001
Sex			
Male (reference)	0.34	N/A	N/A
Female	0.38	1.19 (0.96-1.47)	.11
Race/ethnicity			
White non-Hispanic (reference)	0.44	N/A	N/A
Black non-Hispanic	0.31	0.55 (0.33-0.92)	.002
Hispanic	0.29	0.52 (0.34-0.79)	<.001
Other/unknown	0.42	0.93 (0.58-1.50)	.76
Dual Medicare/Medicaid enrollment (2019)			
Nondual Medicare/Medicaid enrollment (reference)	0.48	N/A	N/A
Any enrollment Medicare/Medicaid	0.26	0.39 (0.26-0.59)	<.001
Census region			
Midwest (reference)	0.36	N/A	N/A
Northeast	0.34	0.93 (0.65-1.32)	.66
South	0.34	0.94 (0.71-1.24)	.65
West	0.42	1.33 (0.94-1.89)	.11
Comorbidities			
0 (reference)	0.39	N/A	N/A
1	0.37	0.90 (0.68-1.19)	.46
≥2	0.34	0.79 (0.61-1.03)	.08

[a]aOR: adjusted odds ratio.

[b]Metropolitan statistical area residence defined by the Office of Management and Budget as having at least one urbanized area with a minimum population of 50,000.

[c]N/A: not applicable.

Telehealth Availability

Older age, rural residence, and dual Medicare/Medicaid enrollment were associated with lower probabilities of telehealth availability (Table 5). Compared to urban survivors, rural survivors had lower predicted probability of telehealth availability (53% vs 63%; *P*<.001). Telehealth availability was not associated with race/ethnicity. Compared to nondual enrollees, dual Medicare/Medicaid-enrolled beneficiaries had lower probability of being offered telehealth (53% vs 63%; *P*=.009).

Table 5. Factors associated with telehealth availability among Medicare Current Beneficiary Survey respondents with a history of cancer and a usual source of care (N=2044).

Characteristic	Telehealth availability		
	Predicted probability	aOR[a] (95% CI)	*P* value
Metropolitan statistical area residence[b]			
Metro (reference)	0.63	N/A[c]	N/A
Nonmetro	0.53	0.68 (0.53-0.87)	<.001
Age group (years)			
65-74 (reference)	0.60	N/A	N/A
<65	0.64	1.18 (0.78-1.79)	.40
≥75	0.50	0.67 (0.54-0.83)	<.001
Sex			
Male (reference)	0.59	N/A	N/A
Female	0.58	0.94 (0.74-1.21)	.65
Race/ethnicity			
White non-Hispanic (reference)	0.59	N/A	N/A
Black non-Hispanic	0.56	0.89 (0.52-1.50)	.65
Hispanic	0.59	1.01 (0.67-1.52)	.95
Other/unknown	0.60	1.06 (0.63-1.77)	.83
Dual Medicare/Medicaid enrollment (2019)			
Nondual Medicare/Medicaid enrollment (reference)	0.63	N/A	N/A
Any enrollment Medicare/Medicaid	0.53	0.66 (0.48-0.90)	.009
Census region			
Midwest (reference)	0.56	N/A	N/A
Northeast	0.56	1.01 (0.74-1.38)	.96
South	0.51	0.86 (0.65-1.13)	.27
West	0.70	1.88 (1.27-2.79)	.002
Comorbidities			
0 (reference)	0.56	N/A	N/A
1	0.58	1.07 (0.83-1.39)	.60
≥2	0.61	1.26 (0.95-1.69)	.11

[a]aOR: adjusted odds ratio.

[b]Metropolitan statistical area residence defined by the Office of Management and Budget as having at least one urbanized area with a minimum population of 50,000.

[c]N/A: not applicable.

Discussion

Principal Findings

Early during the COVID-19 pandemic, we found that over 80% of Medicare-enrolled cancer survivors owned the necessary technology for telehealth encounters, but only half had experience using the internet for communication; almost two-thirds of survivors reported that their usual provider offered telehealth. Consistent with previous research [12-14], study findings highlight a complex digital divide related to telehealth availability and technology ownership and use, particularly among older, Black, Hispanic, lower-income, and rural cancer survivors. Despite the potential of telehealth to meet the unique health care needs of cancer survivors (eg, surveillance, comorbidities, and primary and survivorship care), some patient groups face greater barriers to technology access. These patterned differences in use and access underscore a need to engage multilevel interventions to mitigate the underlying barriers to telehealth use. These results have implications for clinicians, patient advocates, and policy makers as they seek to improve access to care for vulnerable cancer survivors, particularly as the COVID-19 pandemic continues and telehealth

becomes an increasingly important bridge between patients and providers.

Study findings highlight gaps in reported telehealth availability, raising concerns that some providers may have limited telehealth infrastructure [15]. Given the increasingly important role of telehealth to access services, clinicians may need to enhance their practices' telehealth capabilities and clinical workflow by providing additional staff support during video log-on and follow-up processes. New procedures may be needed to assess and refer patients to community resources that can augment technology access and telehealth literacy [16]. Patient advocates and policy makers can support clinician efforts to engage patients via telehealth through continued reimbursement of telehealth visits that support technical and staff requirements. Legislation supporting reimbursement of telehealth services beyond the pandemic, including audio-only telehealth visits, is important in providing equitable access to care among older, rural cancer survivors living in poverty [17].

Gaps in patient access to technology need to be considered within the broader context of structural inequities and policies to address them. Data suggest limited broadband internet access is more prevalent among rural residents, adults 65 years and older, minority populations, and communities of lower socioeconomic status, the same populations that experience disparities in access to cancer survivorship care [18-20]. In parallel with supporting clinicians, ongoing efforts by policy makers to expand broadband access will be essential to reducing disparities in telehealth access. Findings suggest systemic factors may be influencing technology ownership, internet access and use for communication, and telehealth availability, necessitating further monitoring of telehealth in health care delivery to ensure that existing inequities in survivorship care are not exacerbated as telehealth availability increases.

Limitations and Future Research

The cross-sectional data are from June and July 2020, and do not reflect the population's telehealth experience over the course of the COVID-19 pandemic. Technology access and use and provider telehealth availability were self-reported and may be subject to bias; the survey did not measure telehealth use or difficulties with technology. The analytic sample included individuals aged <65 years, a special group of younger patients who are either disabled or have ESRD, and may not be representative of other adult cancer survivors aged <65 years. Further, although sample respondents reported a cancer history, we cannot ascertain if cancer was an active health problem. Despite these limitations, the study draws on a large, population-based sample of cancer survivors, with timely and targeted questions addressing telehealth use early during the COVID-19 pandemic. Future research is needed to assess cancer survivors' experience with telehealth use as the pandemic continues. In addition, research should examine providers' experience with offering their patients telehealth and challenges in serving historically vulnerable populations of cancer survivors. More research is needed to understand the telehealth needs and preferences of Medicare-enrolled cancer survivors, particularly those facing barriers to accessing and using technology necessary for telehealth.

Conclusion

This study captures disparities in telehealth availability and related technological requirements during the COVID-19 pandemic among Medicare-enrolled cancer survivors. Developing and testing multilevel solutions for the "double-burden" of lack of technological access and disparities in access to health care are important to ensure existing inequities in survivorship care are not exacerbated as telehealth becomes more embedded into postpandemic health care delivery models.

Acknowledgments

The opinions expressed by the authors are their own, and this material should not be interpreted as representing the official viewpoint of the US Department of Health and Human Services, the National Institutes of Health, or the National Cancer Institute.

Conflicts of Interest

AJD received consulting fees from Amgen Inc and collaborated with researchers at Genentech and Flatiron, Inc; a close family member received fees for participation on advisory boards for Abbvie and Celgene. None of these were related to the current research. The other authors have no conflicts to disclose.

References

1. Koma W, Cubanski J, Neuman T. Medicare and telehealth: coverage and use during the COVID-19 pandemic and options for the future. KFF. 2021. URL: https://www.kff.org/medicare/issue-brief/medicare-and-telehealth-coverage-and-use-during-the-covid-19-pandemic-and-options-for-the-future/ [accessed 2021-06-23]
2. Koonin LM, Hoots B, Tsang CA, Leroy Z, Farris K, Jolly T, et al. Trends in the use of telehealth during the emergence of the COVID-19 pandemic - United States, January-March 2020. MMWR Morb Mortal Wkly Rep 2020 Oct 30;69(43):1595-1599. [doi: 10.15585/mmwr.mm6943a3] [Medline: 33119561]
3. Medicare telemedicine health care provider fact sheet. Centers for Medicare and Medicaid Services. 2020. URL: https://www.cms.gov/newsroom/fact-sheets/medicare-telemedicine-health-care-provider-fact-sheet [accessed 2021-11-30]

4. Kuderer NM, Choueiri TK, Shah DP, Shyr Y, Rubinstein SM, Rivera DR, COVID-19 and Cancer Consortium. Clinical impact of COVID-19 on patients with cancer (CCC19): a cohort study. Lancet 2020 Jun 20;395(10241):1907-1918 [FREE Full text] [doi: 10.1016/S0140-6736(20)31187-9] [Medline: 32473681]
5. Paterson C, Bacon R, Dwyer R, Morrison KS, Toohey K, O'Dea A, et al. The role of telehealth during the COVID-19 pandemic across the interdisciplinary cancer team: implications for practice. Semin Oncol Nurs 2020 Dec;36(6):151090 [FREE Full text] [doi: 10.1016/j.soncn.2020.151090] [Medline: 33218886]
6. Perrin A, Atske S. 7% of Americans don't use the internet. Who are they? Pew Research Center. 2021. URL: https://www.pewresearch.org/fact-tank/2021/04/02/7-of-americans-dont-use-the-internet-who-are-they/ [accessed 2021-06-23]
7. Bashshur RL, Doarn CR, Frenk JM, Kvedar JC, Shannon GW, Woolliscroft JO. Beyond the COVID pandemic, telemedicine, and health care. Telemed J E Health 2020 Nov;26(11):1310-1313. [doi: 10.1089/tmj.2020.0328] [Medline: 32809913]
8. Hirko K, Kerver J, Ford S, Szafranski C, Beckett J, Kitchen C, et al. Telehealth in response to the COVID-19 pandemic: implications for rural health disparities. J Am Med Inform Assoc 2020 Nov 01;27(11):1816-1818 [FREE Full text] [doi: 10.1093/jamia/ocaa156] [Medline: 32589735]
9. Scott Kruse C, Karem P, Shifflett K, Vegi L, Ravi K, Brooks M. Evaluating barriers to adopting telemedicine worldwide: a systematic review. J Telemed Telecare 2018 Jan;24(1):4-12 [FREE Full text] [doi: 10.1177/1357633X16674087] [Medline: 29320966]
10. Original Medicare (part A and B) eligibility and enrollment. Centers for Medicare and Medicaid Services. 2021. URL: https://www.cms.gov/Medicare/Eligibility-and-Enrollment/OrigMedicarePartABEligEnrol [accessed 2021-11-30]
11. Dual eligible beneficiaries: an overview. MedPAC. URL: https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/June04_ch3.pdf [accessed 2021-09-24]
12. Roberts ET, Mehrotra A. Assessment of disparities in digital access among Medicare beneficiaries and implications for telemedicine. JAMA Intern Med 2020 Oct 01;180(10):1386-1389 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2666] [Medline: 32744601]
13. Lam K, Lu AD, Shi Y, Covinsky KE. Assessing telemedicine unreadiness among older adults in the United States during the COVID-19 pandemic. JAMA Intern Med 2020 Oct 01;180(10):1389-1391 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2671] [Medline: 32744593]
14. Patel SY, Mehrotra A, Huskamp HA, Uscher-Pines L, Ganguli I, Barnett ML. Variation in telemedicine use and outpatient care during the COVID-19 pandemic in the United States. Health Aff (Millwood) 2021 Feb;40(2):349-358. [doi: 10.1377/hlthaff.2020.01786] [Medline: 33523745]
15. Demeke HB, Merali S, Marks S, Pao LZ, Romero L, Sandhu P, et al. Trends in use of telehealth among health centers during the COVID-19 pandemic - United States, June 26-November 6, 2020. MMWR Morb Mortal Wkly Rep 2021 Feb 19;70(7):240-244. [doi: 10.15585/mmwr.mm7007a3] [Medline: 33600385]
16. Zon RT, Kennedy EB, Adelson K, Blau S, Dickson N, Gill D, et al. Telehealth in oncology: ASCO standards and practice recommendations. JCO Oncol Pract 2021 Sep;17(9):546-564. [doi: 10.1200/OP.21.00438] [Medline: 34319760]
17. Noguchi Y. Voice-only telehealth may go away with pandemic rules expiring. NPR. 2021. URL: https://www.npr.org/sections/health-shots/2021/11/23/1056612250/voice-only-telehealth-might-go-away-with-pandemic-rules-set-to-expire [accessed 2021-11-30]
18. DeGuzman PB, Bernacchi V, Cupp CA, Dunn B, Ghamandi BJF, Hinton ID, et al. Beyond broadband: digital inclusion as a driver of inequities in access to rural cancer care. J Cancer Surviv 2020 Oct;14(5):643-652. [doi: 10.1007/s11764-020-00874-y] [Medline: 32390103]
19. FCC Annual Broadband Report shows digital divide is rapidly closing. Federal Communications Commission. 2021. URL: https://www.fcc.gov/document/fcc-annual-broadband-report-shows-digital-divide-rapidly-closing [accessed 2021-06-23]
20. Yoon H, Jang Y, Vaughan PW, Garcia M. Older adults' internet use for health information: digital divide by race/ethnicity and socioeconomic status. J Appl Gerontol 2020 Jan;39(1):105-110. [doi: 10.1177/0733464818770772] [Medline: 29661052]

Abbreviations

ESRD: end-stage renal disease

Edited by C Basch; submitted 02.11.21; peer-reviewed by S Okoye, H Zheng, A Billis; comments to author 20.11.21; revised version received 17.12.21; accepted 19.12.21; published 25.01.22.

Please cite as:
Lama Y, Davidoff AJ, Vanderpool RC, Jensen RE
Telehealth Availability and Use of Related Technologies Among Medicare-Enrolled Cancer Survivors: Cross-sectional Findings From the Onset of the COVID-19 Pandemic
J Med Internet Res 2022;24(1):e34616
URL: https://www.jmir.org/2022/1/e34616
doi:10.2196/34616
PMID:34978531

Original Paper

Challenges of Telemonitoring Programs for Complex Chronic Conditions: Randomized Controlled Trial With an Embedded Qualitative Study

Patrick Ware[1*], MPH, PhD; Amika Shah[1,2*], BSc, MScCH; Heather Joan Ross[3,4], MHSc, MD; Alexander Gordon Logan[2,5,6,7], MD; Phillip Segal[8,9], MD; Joseph Antony Cafazzo[1,2,10], PEng, PhD; Katarzyna Szacun-Shimizu[1], HBA, MA; Myles Resnick[1], BSc, MSc; Tessy Vattaparambil[1], BSc, MHI; Emily Seto[1,2], PEng, PhD

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

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

[3]Peter Munk Cardiac Centre, University Health Network, Toronto, ON, Canada

[4]Ted Rogers Centre for Heart Research, University Health Network, Toronto, ON, Canada

[5]Lunenfeld-Tanenbaum Research Institute, Mount Sinai Hospital, Toronto, ON, Canada

[6]Division of Nephrology, Mount Sinai Hospital, Toronto, ON, Canada

[7]Division of Nephrology, University Health Network, Toronto, ON, Canada

[8]Division of Endocrinology, University Health Network, Toronto, ON, Canada

[9]Department of Medicine, University of Toronto, Toronto, ON, Canada

[10]Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada

[*]these authors contributed equally

Corresponding Author:
Patrick Ware, MPH, PhD
Centre for Global eHealth Innovation
University Health Network
190 Elizabeth St.
Toronto, ON, M5G 2C4
Canada
Phone: 1 (416) 340 4800 ext 4765
Email: Patrick.Ware@uhn.ca

Abstract

Background: Despite the growing prevalence of people with complex conditions and evidence of the positive impact of telemonitoring for single conditions, little research exists on telemonitoring for this population.

Objective: This randomized controlled trial and embedded qualitative study aims to evaluate the impact on and experiences of patients and health care providers (HCPs) using a telemonitoring system with decision support to manage patients with complex conditions, including those with multiple chronic conditions, compared with the standard of care.

Methods: A pragmatic, unblinded, 6-month randomized controlled trial sought to recruit 146 patients with ≥1 diagnosis of heart failure (HF), uncontrolled hypertension (HT), and insulin-requiring diabetes mellitus (DM) from outpatient specialty settings in Toronto, Ontario, Canada. Participants were randomized into the control and telemonitoring groups, with the latter being instructed to take readings relevant to their conditions. The telemonitoring system contained an algorithm that generated decision support in the form of actionable self-care directives to patients and alerts to HCPs. The primary outcome was health status (36-Item Short Form Health Survey questionnaire). Secondary outcomes included anxiety and depression, self-efficacy in chronic disease management, and self-reported health service use. HF-related quality of life and self-care measures were also collected from patients followed for HF. Within- and between-group change scores were analyzed for statistical significance (P<.05). A convenience sample of HCPs and patients in the intervention group was interviewed about their experiences.

Results: A total of 96 patients were recruited and randomized. Recruitment was terminated early because of implementation challenges and the onset of the COVID-19 pandemic. No significant within- and between-group differences were found for the main primary and secondary outcomes. However, a within-group analysis of patients with HF found improvements in self-care maintenance (P=.04) and physical quality of life (P=.046). Opinions expressed by the 5 HCPs and 13 patients who were interviewed

differed based on the monitored conditions. Although patients with HF reported benefitting from actionable self-care guidance and meaningful interactions with their HCPs, patient and HCP users of the DM and HT modules did not think telemonitoring improved the clinical management of those conditions to the same degree. These differing experiences were largely attributed to the siloed nature of specialty care and the design of the decision support, whereby fluctuations in the status of HT and DM typically required less urgent interventions compared with patients with HF.

Conclusions: We recommend that future research conceive telemonitoring as a *program* and that self-management and clinical decision support are necessary but not sufficient components of such programs for patients with complex conditions and lower acuity. We conclude that telemonitoring for patients with complex conditions or within multidisciplinary care settings may be best operationalized through nurse-led models of care.

Trial Registration: ClinicalTrials.gov NCT03127852; https://clinicaltrials.gov/ct2/show/NCT03127852

International Registered Report Identifier (IRRID): RR2-10.2196/resprot.8367

(J Med Internet Res 2022;24(1):e31754) doi:10.2196/31754

KEYWORDS

telemonitoring; telemedicine; heart failure; diabetes; hypertension; tertiary health care; multiple chronic conditions; mobile phone

Introduction

Telemonitoring for the Management of Chronic Conditions

Despite a growing prevalence of patients with complex and multiple chronic conditions (MCCs) [1,2], siloed care models focusing on single conditions have been a barrier to the appropriate management of these patients [3]. In Canada, 12.9% of individuals across all age groups report having ≥2 chronic conditions, and 3.9% report having ≥3 conditions [1]. They are among the highest cost users of health care systems because of a higher frequency of hospitalizations, many of which are thought to be preventable [4,5]. Research suggests that effective patient self-management can reduce the need for urgent care while promoting self-efficacy, improving quality of life, and reducing the risk of adverse psychological effects [6]. However, complex decision-making and often conflicting clinical advice from multiple siloed health care providers (HCPs) make self-management challenging for patients with MCCs [7,8].

Telemonitoring has the potential to empower patients and HCPs by facilitating patient self-management and clinical decision support to manage MCCs. Telemonitoring systems enable patients to track vital signs and symptoms and can enable the automatic generation of self-management instructions [9]. In addition, by delivering these data to the clinical team, HCPs can identify patients showing early signs of exacerbation, which offers an opportunity for reinforcing the principles of self-management at *teachable moments* [10]. Importantly, telemonitoring allows HCPs to provide remote guidance or make changes to a care plan, thereby stabilizing symptoms before they escalate to the point of hospitalization.

Although research on telemonitoring is rapidly growing [11], these systems typically target a singular condition such as diabetes mellitus (DM), hypertension (HT), or heart failure (HF) [12-14]. Systematic reviews indicate that telemonitoring for single conditions leads to improved health outcomes and quality of life and reductions in health service use and costs [15-21]. Studies that do not report improvements do not often include a self-care component, are difficult to use, or do not target patients who are most ill and frequently hospitalized [22-25]. To date, few studies have focused on the use of telemonitoring among patients with complex conditions, although these patients may benefit the most from such interventions [15,26-32].

Objective

The objective of this study is to evaluate the impact of a mobile phone–based telemonitoring program for the management of patients with complex conditions in specialty care settings. Patients with complex conditions are defined as those who are at high risk for hospitalization, exacerbations of their chronic conditions, and disease progression and those with MCCs, including HF, uncontrolled HT, and insulin-requiring DM. The primary research question was the following: *what is the impact of a telemonitoring program for patients with complex conditions on health status, self-management, and health service use?* The secondary research question was the following: *what were the experiences of patients and HCPs related to the telemonitoring program and the way in which it was implemented?*

Methods

Study Design and Setting

This was a pragmatic, unblinded, 1:1 randomized controlled trial (RCT) comparing the 6-month impact of telemonitoring to support the management of patients with complex conditions with that of the standard of care. An embedded qualitative component was included to understand the results of the trial. Patients with HF were recruited from the Heart Function Clinic at the University Health Network (UHN), a large academic hospital in Toronto, Ontario, Canada, between August 2016 and February 2018. Patients with HT were recruited from an HT clinic at Mount Sinai Hospital between July 2019 and December 2019, and patients with DM were recruited from the UHN Endocrinology Clinic between August 2019 and December 2019. The study received approval from the UHN (15-9995-BE) and Mount Sinai Hospital (16-0093-E) research ethics boards, which approved the procedures to ensure patient privacy, including anonymization of data during storage and analysis.

Participants

To be eligible, patients had to be aged ≥18 years; able to speak and read English (or have a caregiver who does); and diagnosed with HF with reduced ejection fraction (<40%), uncontrolled HT (≥140/90 mm Hg auscultatory), or insulin-requiring DM and performing self-capillary glucose monitoring. Exclusion criteria included being on mechanical circulatory support, dialysis, or a transplant list. In addition, patients with a life expectancy <1 year, dementia, uncontrolled psychiatric illness, or residents of a long-term care facility were excluded. Refer to the published protocol for the full criteria [33].

Treatment Arms

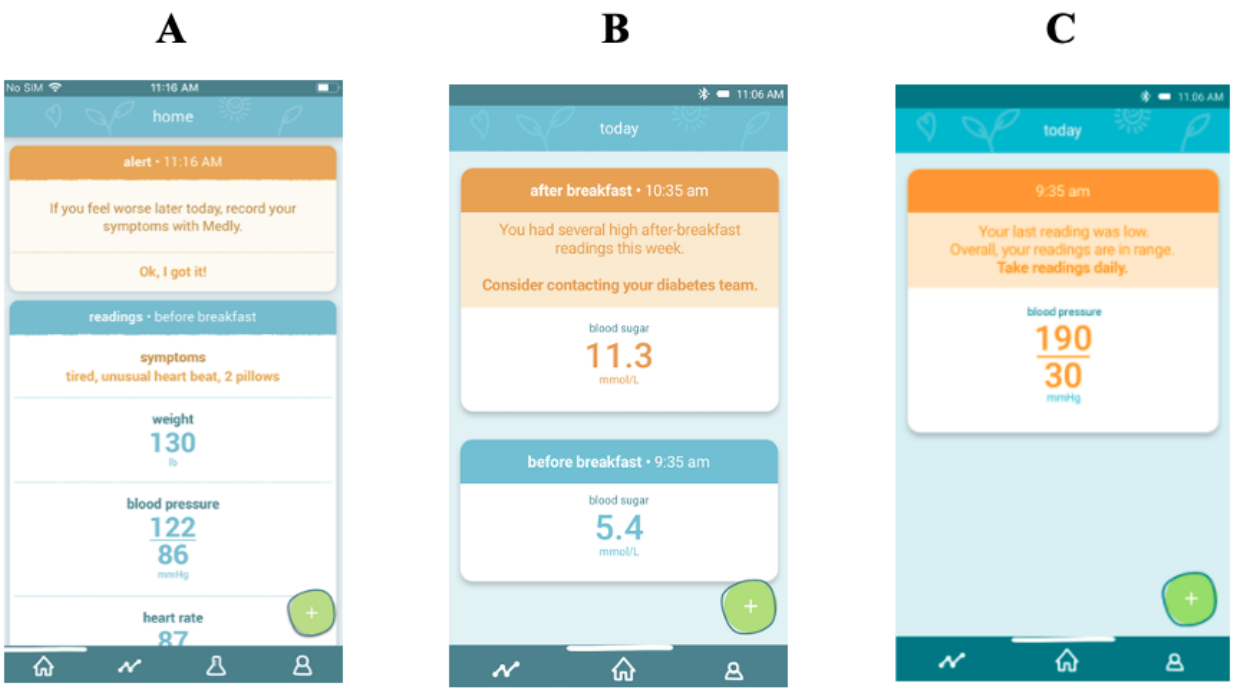
Telemonitoring Group

The intervention was a mobile phone–based telemonitoring program involving a system named Medly (UHN), which enables patients with chronic conditions to take relevant physiological measurements with wireless home medical devices and answer symptom questions using the Medly smartphone app. In response to these inputs, rule-based algorithms, which were iteratively developed and validated by HF, DM, and HT specialists [34] and customized through target thresholds, displayed self-care instructions to patients (Figure 1) and sent alerts to the clinical team via email and a secure web portal where historical trends could also be viewed [34]. As such, the system was designed to improve patient self-management and provide clinical decision support to HCPs [35]. Participants were provided with the necessary equipment, including a smartphone and relevant Bluetooth devices (weight scale, blood pressure monitor, and blood glucose monitor) [33].

Patients with HF were instructed to monitor their daily weight, blood pressure, heart rate, and symptoms. Owing to the frequency of readings and higher complexity of the HF Medly algorithm [36], patient feedback was designed to be highly actionable. For example, patients were told to take a dose of their prescribed diuretic and restrict salts and fluids upon a large weight gain. Patients with DM were instructed to record blood glucose readings at least once per week or more as instructed by their HCP and received actionable feedback (eg, “Eat 15 g [1 tbls] of sugar or other fast-acting carbohydrates” in response to low blood glucose readings). Patients with HT were instructed to report their blood pressure once every 2 weeks unless the readings were out of range, in which case, they would be instructed through the app to increase the frequency of their readings. All modules had messages instructing the users to contact the clinic or go to the emergency department (ED) when critical parameters were out of range. To assist with adherence, an automated phone call was sent to patients based on the required frequency of each condition’s algorithm.

Although the intent was for patients with MCCs to be followed holistically, development delays for the HT and DM modules led to patients with HF being enrolled and managed for HF alone, even if they had MCCs. In addition, the siloed nature of specialty care made it such that even when the technology could support the simultaneous management of HF, DM, and HT, patients were only monitored for the condition being managed at the location of enrollment. As a result, the model of care differed depending on the structure and resources at each clinic. For example, alerts for patients with HF were primarily addressed by nurse practitioners who would escalate issues to the treating cardiologist as required. In contrast, telemonitoring alert management for HT and DM was the responsibility of the treating physician.

Figure 1. Screens of the Medly multiple chronic conditions app showing self-care feedback for (A) heart failure, (B) diabetes mellitus, and (C) hypertension.



Standard of Care Group

Standard of care followed Canadian clinical care guidelines for HF, DM, and HT [37-39]. In general, that included seeing the clinical team for scheduled follow-ups every 3 to 6 months, optimization of medical therapy, and self-management education.

Enrollment and Randomization

A total of 146 patients with varying chronic conditions were targeted for enrollment (see sample size justification [33]). HCPs familiar with the study's inclusion criteria identified patients during scheduled outpatient appointments and introduced them to an on-site research coordinator. After confirming eligibility, the coordinator explained the study and obtained informed consent. The patients were then block randomized using blocks of 4 into the intervention and standard of care treatment groups in a 1:1 ratio as per the published stratification protocol [33]. Patients allocated to the intervention arm were provided with the telemonitoring equipment and user manual before receiving face-to-face training on how to use the equipment.

Outcomes

The primary outcome was health status as measured by the 36-Item Short Form Health Survey questionnaire. Secondary outcomes included anxiety and depression, as measured using the Hospital Anxiety and Depression Scale [40], and patients' self-efficacy to manage their condition, as measured using the Self-Efficacy for Managing Chronic Disease 6-Item scale [41-43]; the number of self-reported interactions with the health system in the previous 6 months was also collected, including hospitalizations and visits to ED, specialty care clinics, and family physicians. Self-care, as measured by the Self-Care of HF Index [44], and HF-specific quality of life, as measured by the Minnesota Living with HF Questionnaire, [45] were also collected for patients with HF.

Quantitative Data Collection and Analysis

Questionnaires containing the patient-reported outcome measures were administered at baseline and at 6 months [33]. Demographic questions were included in the baseline questionnaire to characterize the study participants. Patient adherence was calculated from the Medly server log data as a percentage of the completed recommended readings over the course of the 6-month trial.

Posttrial data and change scores were compared between the treatment arms using independent Student *t* tests and Mann–Whitney tests (for normally and not normally distributed data, respectively). Paired Student *t* tests and Wilcoxon signed-rank tests were performed to compare baseline and poststudy data within the control and telemonitoring groups. Analyses were performed using SPSS (version 27; IBM Corp) under the intention-to-treat principle and using a significance of $P<.05$.

Qualitative Data Collection and Analysis

All HCPs who used the system and a sample of patients in the intervention group were invited to participate in poststudy semistructured interviews. The interviews began with open-ended questions about the participants' experiences with the intervention and had probing questions inspired by the constructs of performance expectancy, effort expectancy, social influence, and facilitating conditions from the unified theory of acceptance and use of technology [46]. Patients were identified through a convenience sample with efforts made to include patients from the 3 monitored conditions. Interviews took place in a private clinic room or over the telephone and were audio recorded for later transcription.

Qualitative data were analyzed using a conventional content analysis approach [47] by 2 researchers (AS and PW). An initial round of independent open coding was conducted with the primary objective of organizing quotes into themes that explained the quantitative results. Then, AS and PW met to discuss themes and agree upon a coding framework that was applied in the second round of deductive coding. A final discussion was held to review the results and reach a consensus. NVivo (QSR International; version 11) was used to help organize the source documents and themes.

Results

Study Participants

Recruitment ended after the enrollment of 96 patients, of which 66 (69%) were followed for HF, 22 (23%) were followed for uncontrolled HT, 5 (5%) were followed for insulin-requiring DM, and 3 (3%) were followed for both HT and DM (Figure 2). The decision to stop recruitment before reaching the target sample size was made for 3 key reasons. First, the HF telemonitoring program became standard of care at the UHN Heart Function clinic, meaning there would no longer be a difference between the treatment arms. Second, recruitment challenges were observed for patients with DM because of the rapid emergence and growing use of continuous and flash glucose monitors, which meant that few patients met the criteria of self-capillary glucose monitoring. Finally, and most importantly, the COVID-19 pandemic led to a pause of nonessential research activities and a significant shift toward virtual care, which fundamentally altered the control group after the research pause was lifted.

In addition to ending recruitment, shifting priorities at the onset of the COVID-19 pandemic also affected the collection of poststudy data, as patients followed for HT and DM ended their enrollment during the first wave, which led to a higher rate of incomplete questionnaires. A death was reported in each treatment arm; however, these were not considered as adverse events of the study. Therefore, although 45 and 50 patients were allocated to the telemonitoring and control arms, only 29% (28/96) and 27% (26/96) of complete data sets were available for analysis in the telemonitoring and control arms, respectively. Study participants were predominantly male (54/96, 56%) and had an average age of 59 (SD 12.6) years. These, along with the other demographics presented in Table 1, are representative of the UHN Heart Function Clinic, from which most patient participants were recruited.

Figure 2. Flow of patient participants through the trial.

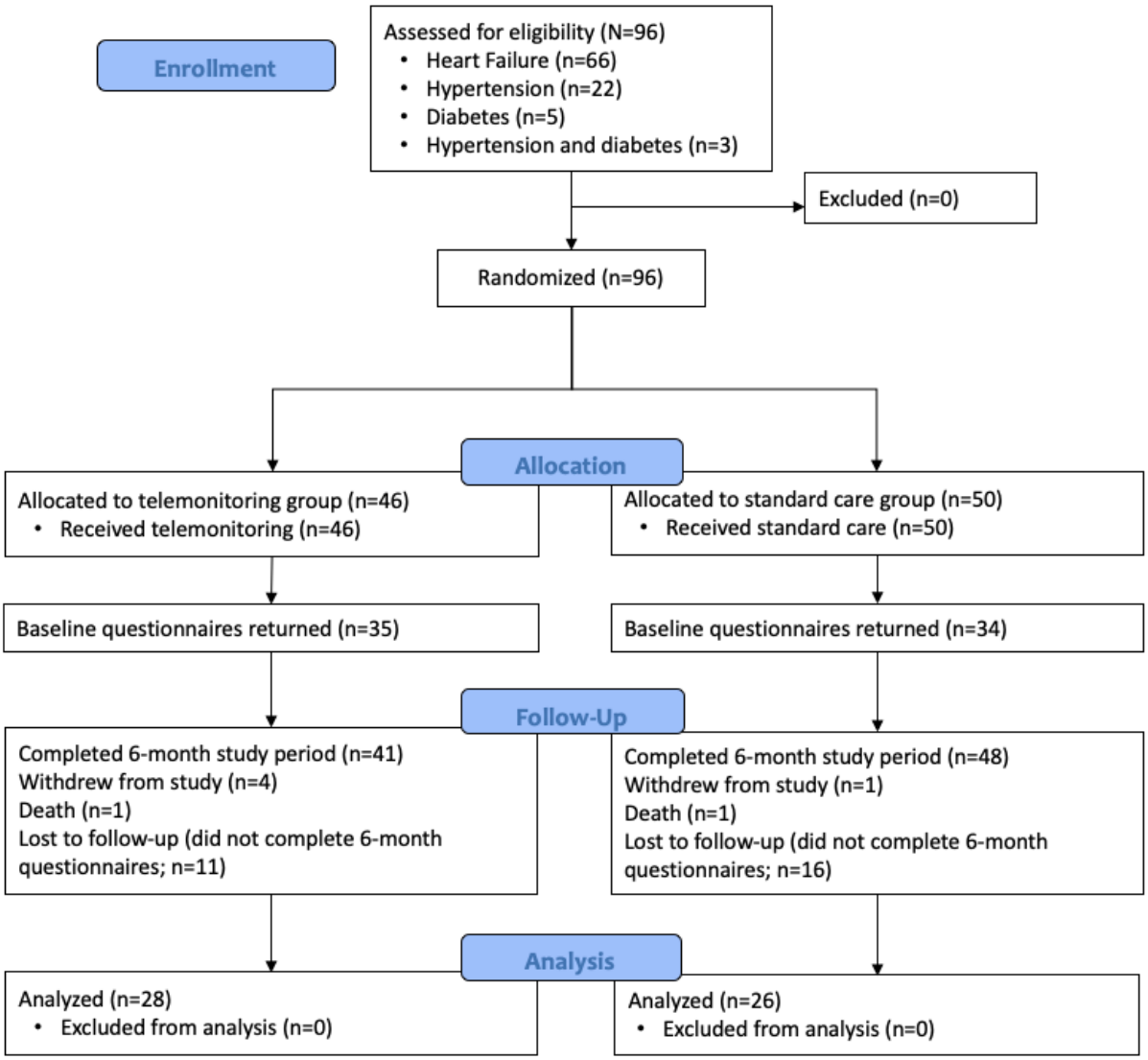


Table 1. Baseline characteristics of patient participants (N=96).

Characteristics	Telemonitoring group (n=46)	Control group (n=50)	Total
Age (years), mean (SD)	62 (12.6)	55 (14.3)	59 (13.9)
Sex, n (%)			
Male	28 (61)	26 (52)	54 (56)
Female	9 (20)	8 (16)	17 (18)
Ethnicity, n (%)			
White	22 (48)	24 (48)	46 (48)
Black	2 (4)	4 (8)	6 (6)
Asian	3 (7)	6 (12)	9 (9)
Other	7 (15)	2 (4)	9 (9)
Rurality, n (%)			
Urban	20 (43)	23 (46)	43 (45)
Suburban	11 (24)	11 (22)	22 (23)
Rural	2 (4)	2 (4)	2 (2)
Highest education achieved, n (%)			
Less than high school	4 (9)	1 (2)	5 (5)
High school	5 (11)	4 (8)	9 (9)
Trade or technical training	5 (11)	6 (12)	11 (11)
College or university	15 (33)	19 (38)	34 (35)
Postgraduate	5 (11)	5 (10)	10 (10)
Comfort with smartphone, n (%)			
Not comfortable	1 (2)	0 (0)	1 (1)
Somewhat comfortable	8 (17)	4 (8)	12 (13)
Comfortable	8 (17)	28 (56)	16 (17)
Very comfortable	13 (28)	15 (30)	28 (29)

Quantitative Outcomes

Table 2 shows the results from the statistical analyses. The within-group pre–post comparisons, poststudy between-group comparisons, and between-group change score comparisons revealed no significant differences in the primary outcome of health status as measured by the 36-Item Short Form Health Survey questionnaire physical and emotional subscales. Similarly, no statistically significant differences were observed for any of the secondary outcomes relevant to the entire sample, including anxiety and depression (as measured by the Hospital Anxiety and Depression Scale), self-efficacy (as measured by the Self-Efficacy for Managing Chronic Disease 6-Item scale), and self-reported use metric. An exception was a reduction in self-reported hospitalizations across all patients, which was significant in the control group.

Table 2. Independent Student *t* test[a] for SF-36[b], HADS[c], SEMCD6[d], self-reported use, MLHFQ[e], and SCHFI[f].

Parameter	Telemonitoring group (n=46)				Standard care group (n=50)				Between-group poststudy data, *P* value	Between-group change scores, *P* value
	Values, n (%)	Baseline, mean (SD)	Poststudy, mean (SD)	*P* value	Values, n (%)	Baseline, mean (SD)	Poststudy, mean (SD)	*P* value		
SF-36										
Physical component	24 (52)	40.94 (8.11)	42.77 (8.58)	.16	25 (50)	41.55 (8.86)	42.39 (9.47)	.24	.48	.63
Mental component	24 (52)	46.42 (12.21)	43.77 (12.28)	.18	25 (50)	46.96 (11.21)	48.31 (10.51)	.53	.73	.37
HADS										
Anxiety	24 (52)	6.94 (4.38)	7.33 (5.07)	.47	22 (44)	6 (4)	5.97 (3.5)	.50	.06	.12
Depression	23 (50)	5.09 (3.95)	5.51 (4.67)	.73	23 (46)	5.21 (3.85)	5.55 (4.2)	.92	.77	.32
SEMCD6	26 (57)	7.23 (2.18)	7.35 (1.57)	.73	24 (48)	7.08 (2.55)	6.75 (2.21)	.53	.13	.42
Self-report										
Hospital (number of visits)	28 (61)	4.29 (10.56)	1.43 (4.11)	.10	23 (46)	3.7 (6.51)	0.35 (1.11)	.02	.02	.32
ED[g] visits	28 (61)	1.29 (4.65)	0.5 (1.33)	.34	23 (46)	0.071 (1.34)	0.22 (0.57)	.07	.12	.31
Clinic visits	20 (44)	3.54 (4.11)	3.71 (2.85)	.52	21 (42)	4.19 (6.01)	2.95 (4.61)	.39	.39	.12
Family physician visits	25 (54)	1.65 (1.22)	1.28 (1.62)	.49	23 (46)	2.48 (3.72)	1.85 (2.23)	.27	.28	.07
SCHFI										
Maintenance	24 (52)	73.42 (14.77)	79.83 (15.81)	.04	17 (34)	73.64 (15.75)	76.18 (17.48)	.49	.74	.82
Management	24 (52)	71.75 (15.64)	75.42 (19.83)	.42	17 (34)	64.47 (27.51)	69.53 (26.53)	.40	.67	.40
Confidence	24 (52)	69.43 (15.49)	71.83 (19.74)	.50	16 (32)	69.94 (25.09)	76.38 (19.86)	.16	.92	.34
MLHFQ										
Total	23 (50)	40.85 (28.52)	35.88 (23.65)	.12	17 (34)	42.81 (26.15)	31.55 (29.45)	.07	.67	.10
Physical	23 (50)	19.13 (11.01)	14.58 (10.9)	.046	17 (34)	20.78 (10.64)	19.05 (12.4)	.11	.43	.06
Emotional	22 (48)	10.38 (7.97)	9 (7.45)	.81	15 (30)	11.38 (7.08)	9.33 (7.1)	.10	.64	.20

[a]A 2-tailed *t* test was used.

[b]SF-36: 36-Item Short Form Health Survey questionnaire.

[c]HADS: Hospital Anxiety and Depression Scale.

[d]SEMCD6: Self-Efficacy for Managing Chronic Disease 6-Item scale.

[e]MLHFQ: Minnesota Living with Heart Failure Questionnaire.

[f]SCHFI: Self-Care of Heart Failure Index.

[g]ED: emergency department.

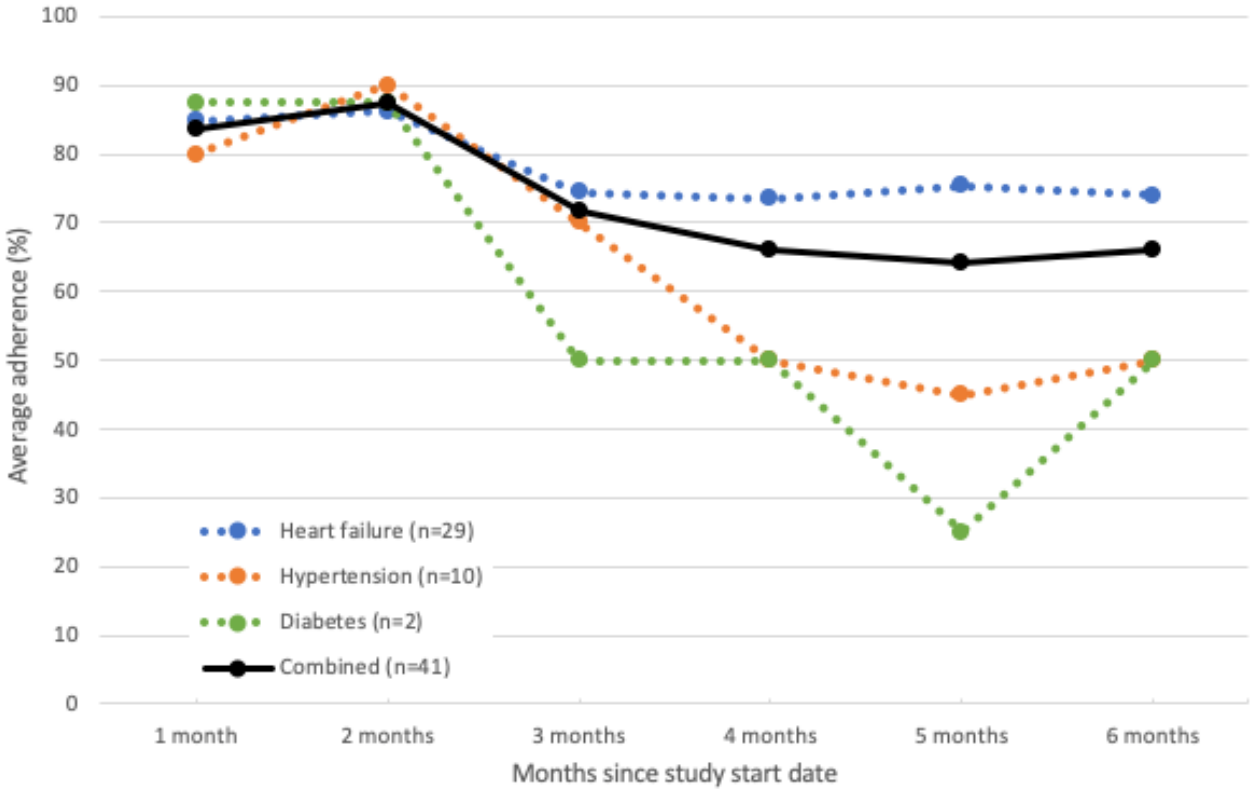
Self-care maintenance and physical quality of life improved significantly for patients with HF in the telemonitoring group (*P*=.04 and *P*=.046, respectively). However, none of the between-group comparisons were statistically significant, likely because of the general improvement in self-care and quality of life scores of the control group and insufficient sample size for detecting changes in condition-specific metrics.

Telemonitoring Use

Of the 41 patients who completed the study in the intervention arm, 29 (71%) used the HF module, 10 (24%) used the HT module, and 2 (5%) used the DM module. On average, patients with HF completed all 4 readings on 78.1% (SD 5.8%) of the days that they were enrolled in the 6-month trial. This use rate increased to 86.2% when looking at the percentage of days the patients with HF reported at least one reading. Average

adherence was 64.2% (SD 18.6%) for the patients with HT and 58.3% (SD 24.6%) for the patients with DM based on a minimum of biweekly and weekly readings for HT and DM, respectively. The combined average adherence across all participants was 73.1% (SD 10%) over the study period. Figure 3 depicts the monthly adherence rates for the 3 disease modules. It shows that although adherence was initially high across all conditions and remained relatively stable for patients with HF, it dropped markedly after the second month for patients with HT and DM.

Figure 3. Average combined and condition-specific adherence rates over time.



Experiences With the Telemonitoring Program

Overview

Of the 41 patients who completed the study in the intervention arm, 13 (32%) were interviewed, including 62% (8/13) of patients with HF, 31% (4/13) of patients with HT, and 8% (1/13) of patients with DM. The HCPs included 1 cardiologist, 2 cardiac nurse practitioners, 1 HT specialist, and 1 endocrinologist. The interview findings, which contributed to explaining the overall null results of the study, are summarized in the following themes: (1) challenges of implementing MCC telemonitoring in a siloed health care system, (2) perceptions of the telemonitoring system, (3) perceived benefits differed based on the condition monitored, and (4) opportunities for MCC telemonitoring.

Theme 1: Challenges of Implementing MCC Telemonitoring in a Siloed Health Care System

Patients and HCPs expressed that the telemonitoring intervention did not adequately address the challenge of delivering and coordinating advanced care for patients with MCCs. Although this was due in part to development delays that resulted in patients with different conditions undergoing the study at different times, there were also no formal communication processes in place among relevant specialties that would have enabled optimal multidisciplinary care:

> *How do you integrate information from multiple specialists...I don't have multiple chronic conditions, I have one condition...I didn't do the blood sugar stuff, I didn't do the heart failure stuff, I didn't do the chronic obstructive pulmonary disease stuff, I didn't do the mental health stuff...The people that would handle all these conditions simultaneously would be family doctors, except family doctors often don't have the expertise to handle complex conditions.* [HCP 5]

As the model of care did not connect existing specialty siloes, the clinical processes surrounding the use of the telemonitoring system were different based on the recruitment setting. For HF, an HCP described that a strong, team-based model with clearly defined clinical roles enabled rapid clinical response to worsening conditions:

> *It's given me different work to do...although [I'm] not complaining about that because that is part of why I'm hired. It's just that...in order for [a telemonitoring program] to work, you have to have a clinician who is devoting time to do all of that, to answer alerts, to document, and to see patients that are unwell in clinic or fax them lab work [requisitions].* [HCP 3]

In contrast, DM and HT alerts were directly sent to the physicians who reflected on the absence of an appropriate model

of care to support telemonitoring within those clinics and how it might have been better if the information was sent to a different member of the team:

> *The difference between the heart failure and the hypertension is that the heart failure has a whole system in place. We don't have a similar type of system for the hypertension, I assume for the diabetes as well...the physician is the last person that needs the information. It's their nurse and dietician.* [HCP 4]

Theme 2: Perceptions of the Telemonitoring System

Most patients had positive opinions about the telemonitoring system and found it easy to take readings and navigate the different features:

> *It's pretty easy to use...I loved the fact that you took your sugar and it connected right away to the phone...I think that for my purposes, it did everything I needed it to do. You got some averages in there too which I liked, you could see [readings] over a period of time.* [Patient with DM 1]

Similarly, patients largely appreciated the automated self-care messages, with the exception of a minority who expressed confusion about messages directing them to contact the clinic when a reading was out of their target range. Some patients, across all conditions, recounted not knowing where to call and what to say when they did. In addition, some patients with HF reported ignoring such messages as they realized that their care team would contact them in the case of a true concern:

> *I think that perhaps the instructions in the system of the Medly program were maybe unclear. Sometimes there would be a prompt to call...[but] then I would call and there wouldn't really be a sort of a good answer on why I should be calling.* [Patient with HT 3]

> *It'll say "contact the heart clinic or go to your hospital if you don't feel well". I've just gotten so used to it. Say my weight is up two pounds or something, I know it's not critical...it says call, but I haven't been...I know that [my care team] will call me and make sure everything's okay.* [Patient with HF 26]

Unlike patients, there were mixed opinions about the telemonitoring system among HCPs. These differences were attributed to the physiological nature of the chronic condition and the existing standard of care in which the intervention was implemented. For example, owing to the dynamic nature of HF and the potential for rapid decompensation that could lead to hospitalization, telemonitoring alerts typically require immediate action from patients or HCPs. Therefore, HF HCPs were motivated to find ways of incorporating the viewing of telemonitoring data and alerts within their existing workflows:

> *I think the issue for clinicians is that they're building this into their day-to-day busy practice so there are things like that I think can be done [to improve the system] but generally I think the dashboard is pretty sophisticated, it's easy to use, [and] incredibly easy to navigate.* [HCP 2]

Conversely, DM and HT HCPs expressed feeling that the system was built on the premise of identifying an acute change (needed for HF) rather than communicating information about longer-term trends. As the system was not perceived to convey DM and HT telemonitoring data to physicians in a way that is consistent with the existing management of those conditions, there was less incentive to incorporate the use of the system within their workflows:

> *I didn't get any information from [the system]...I got no monthly reports about any of the individuals. The only thing I got was an occasional alert if their blood pressure was extraordinarily high...What you want to do is to map the progress of the patient since their blood pressure is not controlled. Even if it's not an alert, think if you had high blood pressure and your blood pressure was not in the alert level, but it was not controlled. Do you want your provider not to have the information?* [HCP 5]

Finally, a factor specific to DM was the fact that the data from the newer, and increasingly prominent, flash glucose monitors could not be sent to the Medly telemonitoring system:

> *To be honest with you, I kind of stopped using it halfway through because...when I mentioned I was doing the study at the last appointment, [my doctor] said, "Well I guess it's irrelevant, based on the new technologies that are out there for checking your glucose."...he made it sound like that the study wasn't happening anymore.* [Patient with DM 1]

Theme 3: Perceived Benefits Differed Based on the Condition Monitored

Opinions of the models of care and telemonitoring system previously described contributed to the perceived benefits of the intervention, and importantly, how these differed for patients with HF, DM, and HT and HCPs. Patients who were followed for HF expressed that the intervention enabled self-management by helping them form a routine of taking daily readings and associating those results with their behavior. The frequency of data collection and urgency-based alerts sent to the clinical team also contributed to the perception of improved clinical management:

> *Every morning when I get up, that's the first thing I do and it kind of gives me an overview of my condition...I wake up and I know I've been having problems breathing and when I check my weight, I say, okay, I have gained weight...three pounds. I say okay and tie [it] back to what I've eaten the night before.* [Patient with HF 8]

The close monitoring offered by the HF telemonitoring system and model of care led to patients with HF benefiting from peace of mind and a closer relationship with their care team:

> *It gives you peace of mind, because...the doctor's seeing it, which is important, because...when the people who are really interested in your well-being*

> *are that far away, it's nice that you've got that system in place, where if something goes wrong you know right away.* [Patient with HF 2]

Patients followed up for HT and DM expressed many of the same self-management benefits as the patients with HF, including the idea that the system encouraged them to take blood pressure and blood glucose readings more regularly than they normally would and contributed to a greater awareness of life factors that could influence those readings:

> *I discovered...triggers of my problem in blood pressure. Mainly, for example, when I have a little constipation, it has an effect on high blood pressure. Or, I had, for example, a little stress that affected me...Medly allowed me to be more aware of the situation. It's a mindfulness program. I became aware of what I was doing.* [Patient with HT 7]

However, unlike patients with HF, patients with HT and DM did not express the same peace of mind and closer therapeutic relationship with their HCPs as a result of telemonitoring. This was explained by the lack of feedback from HCPs and the perception that HCPs were not reviewing the data to the degree that the patients had expected:

> *I didn't really have any sort of feedback from a physician's point of view as to what the Medly information was telling them. That would have been interesting...[and] made the whole experience sort of more sensible. Otherwise, I'm just slapping a cuff on my arm and taking a reading and so if it was more interactive, it might have had more meaning.* [Patient with HT 3]

This perception was confirmed by the HT and DM HCPs, who felt that the intervention did not affect their clinical management of patients because of the fact that the system did not provide health data in a meaningful way:

> *I had virtually no interaction with the study...I wouldn't have even known if patients came for follow-up, whether they were in the study or not. They didn't volunteer the information; I didn't give any information from MCC about the patients. There was no feedback [from the system].* [HCP 5]

Theme 4: Opportunities for MCC Telemonitoring

Despite the challenges with this trial, HCPs spoke of the potential of telemonitoring, if designed appropriately, to help address the existing limitations of virtual care that is becoming the norm since the onset of the COVID-19 pandemic:

> *You need data points, you can't [provide virtual care] in a vacuum and Medly is a system whereby you can get the data points...What telemonitoring is supposed to do is to give a person a profile between clinic visits and if I can't get that through the regular virtual service, this is where the telemonitoring services come in. [However], it's got to be structured, it's got to be done in a systematic way, and it's got to be done so that the patients can submit the information. The benefit for the patient is then they get feedback between clinic visits, or if they don't get feedback between clinic visits, at least the physician at the clinic visits has reliable information in order to ascertain what next steps have to be done to improve management.* [HCP 5]

In addition to the need to present data in a meaningful way, as highlighted in theme 2, there was consensus among HCPs that a telemonitoring system, especially one for MCC, needs to be integrated with the existing organizational information systems:

> *It has to be integrated with a larger system-wide electronic solution...It duplicates work when I have to open a chart, and open a computer program, and transcribe things manually into the chart...Getting alerted is good but [the] communication piece and that documentation piece is [critical].* [HCP 4]

In assessing the value of telemonitoring, HCPs did not disassociate the design of the telemonitoring system from the design of the model of care. For telemonitoring of complex conditions to be feasible, they envisioned a team-based approach in which a nonphysician (eg, nurse, nurse practitioner, or allied health professional) held a central role in the intervention because of a greater alignment of the scope of practice and existing remuneration models:

> *I guess for nurses and dieticians who are employed by a hospital, [the issue of remuneration] is irrelevant because they have a salary, so it could easily be integrated into their current workflow. But for physicians their only real commodity is their time...so if this is going to add to the current work, or displace current remunerated work it's a no-go.* [HCP 4]

> *You could have a nurse practitioner who leads these programs and works with a staff nurse or that kind of helps triage and call patients back and manage that way but then also a physician if there's a really big problem that you have to go to...you could look at different models of care for these programs.* [HCP 2]

Discussion

Principal Findings

In the absence of high-quality evidence on telemonitoring for managing MCCs [48], this study sought to evaluate the impact of a mobile phone–based telemonitoring program to manage patients with complex conditions in specialty care settings. Through a pragmatic RCT, halted before reaching the intended sample size, we observed no effects of the telemonitoring program on health status, anxiety and depression, self-efficacy, and most health service use metrics. Subgroup analyses of patients with HF in the intervention arm found a significant improvement in physical quality of life and self-care maintenance; however, no differences in these groups were observed between the treatment arms. The reduction in self-reported hospitalization found in the control group could also be seen as a trend in the intervention group but did not reach statistical significance, likely because of the small sample

size. This reduction in both arms may be partly attributed to the impact of the HF clinic in stabilizing patients.

Several challenges were encountered during the trial, necessitating deviations from the published protocol, which likely contributed to the null results of the study. Foremost, this trial did not reach its intended sample size of 146. Although the onset of the COVID-19 pandemic and the rapid shift to virtual care for both arms did bring about the decision to stop recruitment permanently, other important challenges contributed to slow recruitment. Importantly, funding and development challenges contributed to the inability to include the planned chronic obstructive pulmonary disease and chronic kidney disease modules. Similarly, although patients who participated in our study had MCCs, we were unable to offer monitoring for all the conditions simultaneously as the HF, HT, and DM modules were not initially available at the same time. Therefore, the MCC model of care was never fully tested as there was never a need to explore communication workflows for multidisciplinary care that are inherent to the care of MCCs. Finally, recruitment of patients from the diabetes clinic was challenging because of the emergence of newer continuous glucose monitoring technologies at the time of recruitment.

Although patient perceptions of the telemonitoring program were largely positive, the perceived benefits and, consequently, use of the system varied across conditions. Indeed, although all patients in this study started with high adherence, these higher levels were only maintained in patients who were followed for HF. These differences may reflect the more actionable nature of the self-care instructions and more frequent clinician alerts from the HF module relative to those provided by the HT and DM modules. HF tends to be more dynamic and requires more parameters to be taken at a higher frequency, which provides the data inputs necessary for a more sophisticated telemonitoring algorithm. Conversely, HT and DM require fewer inputs at a lower frequency, which typically means that less urgency is required when alerts are generated. This does not suggest that HT and DM are not suitable conditions for systems such as Medly, as patients followed up for these conditions did express benefiting from self-management support. Rather, the opinions expressed by the HT and DM HCPs suggest that the way telemonitoring data are presented to HCPs needs to consider the greater importance of long-term trends versus the acute symptom-based alerting needed for optimal HF management. Considering these differences between conditions may increase the perceived relative advantage [46] of telemonitoring to promote comparable use across conditions.

Various factors related to the implementation of the telemonitoring program may also explain the lack of impact observed. We posit that an initially narrow conceptualization of telemonitoring, focused primarily on the technology rather than the model of care, may have contributed to these implementation issues. For instance, relative to HF [49], minimal consideration was given to the model of care associated with the DM and HT modules. Consistent with previous studies, we recommend that future research conceive telemonitoring as a *program* comprising both the *system* and its associated *model of care*. Critically, implementation study designs should be used to assess the feasibility of the program before conducting a trial so that challenges may be identified and addressed [50].

A notable characteristic thought to enhance the feasibility of telemonitoring for patients with complex conditions and MCCs was having a nurse in a central role as it relates to alert management and care coordination. In contrast to the fee-for-service model for physicians, all HCPs in this study noted that the salaried remuneration of nurses offered greater flexibility needed to monitor patients with MCCs through telemonitoring. Indeed**,** when the same MCC telemonitoring system was applied in a different study within a multidisciplinary nurse-led model of care [51], telemonitoring was not only perceived positively by patients with MCCs but was also normalized within their daily life and routines [52], which was a finding that was only observed among the patients with HF and HCPs in this trial, who also benefited from a nurse-led model. What may explain the stark contrast in the results between these studies is the differences in their underlying program theory. Per Kastner et al [53], effective interventions for older adults with MCCs are often founded upon principles of care coordination, disease prioritization, and patient self-management. From this, we conclude that telemonitoring interventions for patients with complex conditions, especially those with MCCs, should (1) be integrated with usual care workflows and complementary clinical services, (2) be supported by a multidisciplinary model, and (3) include ≥1 element of care coordination (eg, referral pathways, case management, and interoperable information systems). Critically, these 3 elements may be best operationalized by having a nurse play a central role in the intervention.

Comparison With Previous Work

The outcomes of this trial are consistent with recent reviews [48,54], indicating a dearth of evidence on the effectiveness of telemedicine interventions for the population with MCC. In 2020, Kraef et al [48] identified only 1 study reporting no impact of telemedicine on self-reported health status and only moderate improvements in disease control measures such as hemoglobin A_{1c} and systolic blood pressure across 7 meta-analyzed studies. Similarly, Pisa et al [54] found no impact of telemedicine interventions in primary care on quality of life and contradictory results on ED visits and hospital admissions across 5 studies [54].

Importantly, trials reporting significant effects on condition-specific outcomes and no implementation challenges often included a role of a nonphysician HCP in the intervention, most commonly a nurse [30,55-57]. Moreover, several authors have attributed frequent and immediate feedback to patient data as important facilitators in promoting the intended use of telemonitoring systems and the success of their trials [30,57,58]. Thus, the results of this trial and the existing literature converge in identifying care coordination as an important component in telemonitoring interventions for MCCs. We suggest that this be considered as a central component of further studies exploring the feasibility and effectiveness of telemonitoring for MCCs.

Limitations

In addition to the recruitment and implementation challenges previously discussed, this study had limitations attributable to the trial's pragmatic nature. First, although the study period was 6 months, the study duration lasted approximately 4 years with no overlap in the periods in which patients with HF and patients with HT or DM were active. This meant that in addition to inhibiting the exploration of multidisciplinary care coordination, ending the recruitment because of the COVID-19 pandemic contributed to a difference in sample size across conditions. Second, self-reported health service use introduces potential challenges because of recall. Third, unintended selection bias is possible as the recruitment relied on busy clinicians identifying eligible patients during clinic hours. We do not have data on patients who might have been eligible but were never identified by their care team or on patients who were approached but did not agree to be introduced to the research coordinator. Finally, the interviews were conducted with a convenience sample of patients, and although all HCPs who were involved in the study in any significant manner were interviewed, the number of 5 HCPs was relatively small.

Conclusions

Decades of research suggest that telemonitoring for self-management and clinical decision support may assist in managing single chronic conditions; however, its effectiveness for the growing number of patients with complex conditions, including those with MCCs, remains unclear. This 6-month pragmatic RCT sought to evaluate the impact on patients and HCPs and their experiences with a telemonitoring program to manage patients with complex conditions compared with those in standard care. No significant within- and between-group differences were found in the primary outcome of health status or the secondary outcomes of self-efficacy, anxiety, depression, and health service use. However, improvements in physical quality of life and self-care maintenance were observed in patients with HF who were enrolled in the telemonitoring program. Qualitative data suggest that these null findings may be because of the fact that not all disease modules were available at the same time, implementation challenges within the siloed specialty care, varying degrees of acuity and urgency across chronic conditions, and an insufficient sample size attributable to the onset of the COVID-19 pandemic. On the basis of the study findings, we caution that self-management and clinical decision support are necessary but not sufficient components of telemonitoring for patients with complex conditions. We conclude that telemonitoring for patients with complex conditions or those within multidisciplinary care settings may be best operationalized through nurse-led models of care and that the complex nature of these interventions makes it such that they are best studied through feasibility or pragmatic study designs before conducting RCTs.

Acknowledgments

The authors wish to thank the trial participants and the health care providers for both their role in facilitating participant recruitment and in the telemonitoring intervention. This study was funded by the Canadian Institutes for Health Research, which had no role in this study.

Conflicts of Interest

ES, HR, and JC are considered inventors of the Medly system under the intellectual property policies of the University Health Network and may benefit from future commercialization of the technology by University Health Network.

Multimedia Appendix 1
CONSORT-eHEALTH Checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1214 KB - jmir_v24i1e31754_app1.pdf]

References

1. Roberts KC, Rao DP, Bennett TL, Loukine L, Jayaraman GC. Prevalence and patterns of chronic disease multimorbidity and associated determinants in Canada. Health Promot Chronic Dis Prev Can 2015 Aug;35(6):87-94 [FREE Full text] [doi: 10.24095/hpcdp.35.6.01] [Medline: 26302227]
2. van Oostrom SH, Gijsen R, Stirbu I, Korevaar JC, Schellevis FG, Picavet HS, et al. Time trends in prevalence of chronic diseases and multimorbidity not only due to aging: data from general practices and health surveys. PLoS One 2016 Aug;11(8):e0160264 [FREE Full text] [doi: 10.1371/journal.pone.0160264] [Medline: 27482903]
3. Moffat K, Mercer SW. Challenges of managing people with multimorbidity in today's healthcare systems. BMC Fam Pract 2015 Oct 14;16(1):129 [FREE Full text] [doi: 10.1186/s12875-015-0344-4] [Medline: 26462820]
4. Zulman DM, Chee CP, Wagner TH, Yoon J, Cohen DM, Holmes TH, et al. Multimorbidity and healthcare utilisation among high-cost patients in the US Veterans Affairs Health Care System. BMJ Open 2015 Apr 16;5(4):e007771 [FREE Full text] [doi: 10.1136/bmjopen-2015-007771] [Medline: 25882486]
5. Wodchis WP, Austin PC, Henry DA. A 3-year study of high-cost users of health care. Can Med Asso J 2016 Feb 16;188(3):182-188 [FREE Full text] [doi: 10.1503/cmaj.150064] [Medline: 26755672]
6. Siminerio LM. The role of technology and the chronic care model. J Diabetes Sci Technol 2010 Mar 01;4(2):470-475 [FREE Full text] [doi: 10.1177/193229681000400229] [Medline: 20307409]

7. Noël PH, Parchman ML, Williams JW, Cornell JE, Shuko L, Zeber JE, et al. The challenges of multimorbidity from the patient perspective. J Gen Intern Med 2007 Dec;22 Suppl 3:419-424 [FREE Full text] [doi: 10.1007/s11606-007-0308-z] [Medline: 18026811]
8. Liddy C, Blazkho V, Mill K. Challenges of self-management when living with multiple chronic conditions: systematic review of the qualitative literature. Can Fam Physician 2014 Dec;60(12):1123-1133 [FREE Full text] [Medline: 25642490]
9. Vassilev I, Rowsell A, Pope C, Kennedy A, O'Cathain A, Salisbury C, et al. Assessing the implementability of telehealth interventions for self-management support: a realist review. Implement Sci 2015;10:59 [FREE Full text] [doi: 10.1186/s13012-015-0238-9] [Medline: 25906822]
10. Seto E, Morita PP, Tomkun J, Lee TM, Ross H, Reid-Haughian C, et al. Implementation of a heart failure telemonitoring system in home care nursing: feasibility study. JMIR Med Inform 2019 Jul 26;7(3):e11722 [FREE Full text] [doi: 10.2196/11722] [Medline: 31350841]
11. Meystre S. The current state of telemonitoring: a comment on the literature. Telemed J E Health 2005 Feb;11(1):63-69. [doi: 10.1089/tmj.2005.11.63] [Medline: 15785222]
12. Logan AG, Irvine MJ, McIsaac WJ, Tisler A, Rossos PG, Easty A, et al. Effect of home blood pressure telemonitoring with self-care support on uncontrolled systolic hypertension in diabetics. Hypertension 2012 Jul;60(1):51-57 [FREE Full text] [doi: 10.1161/HYPERTENSIONAHA.111.188409] [Medline: 22615116]
13. Istepanian RS, Zitouni K, Harry D, Moutosammy N, Sungoor A, Tang B, et al. Evaluation of a mobile phone telemonitoring system for glycaemic control in patients with diabetes. J Telemed Telecare 2009;15(3):125-128. [doi: 10.1258/jtt.2009.003006] [Medline: 19364893]
14. Goldberg LR, Piette JD, Walsh MN, Frank TA, Jaski BE, Smith AL, et al. Randomized trial of a daily electronic home monitoring system in patients with advanced heart failure: the Weight Monitoring in Heart Failure (WHARF) trial. Am Heart J 2003 Oct;146(4):705-712. [doi: 10.1016/S0002-8703(03)00393-4] [Medline: 14564327]
15. Liang HY, Lin LH, Chang CY, Wu FM, Yu S. Effectiveness of a nurse-led tele-homecare program for patients with multiple chronic illnesses and a high risk for readmission: a randomized controlled trial. J Nurs Scholarsh 2021 Mar;53(2):161-170. [doi: 10.1111/jnu.12622] [Medline: 33507626]
16. Inglis SC, Clark RA, McAlister FA, Ball J, Lewinter C, Cullington D, et al. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. Cochrane Database Syst Rev 2010(8):CD007228. [doi: 10.1002/14651858.CD007228.pub2] [Medline: 20687083]
17. Clark RA, Inglis SC, McAlister FA, Cleland JG, Stewart S. Telemonitoring or structured telephone support programmes for patients with chronic heart failure: systematic review and meta-analysis. Br Med J 2007 May 5;334(7600):942 [FREE Full text] [doi: 10.1136/bmj.39156.536968.55] [Medline: 17426062]
18. Polisena J, Tran K, Cimon K, Hutton B, McGill S, Palmer K, et al. Home telemonitoring for congestive heart failure: a systematic review and meta-analysis. J Telemed Telecare 2010;16(2):68-76. [doi: 10.1258/jtt.2009.090406] [Medline: 20008054]
19. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]
20. Bowles KH, Holland DE, Horowitz DA. A comparison of in-person home care, home care with telephone contact and home care with telemonitoring for disease management. J Telemed Telecare 2009;15(7):344-350 [FREE Full text] [doi: 10.1258/jtt.2009.090118] [Medline: 19815903]
21. Hamadi HY, Martinez D, Xu J, Silvera GA, Mallea JM, Hamadi W, et al. Effects of post-discharge telemonitoring on 30-day chronic obstructive pulmonary disease readmissions and mortality. J Telemed Telecare 2020 Nov 11:1357633X20970402. [doi: 10.1177/1357633X20970402] [Medline: 33176540]
22. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, Telemedical Interventional Monitoring in Heart Failure Investigators. Impact of remote telemedical management on mortality and hospitalizations in ambulatory patients with chronic heart failure: the telemedical interventional monitoring in heart failure study. Circulation 2011 May 3;123(17):1873-1880 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.018473] [Medline: 21444883]
23. Koehler F, Winkler S, Schieber M, Sechtem U, Stangl K, Böhm M, et al. Telemedicine in heart failure: pre-specified and exploratory subgroup analyses from the TIM-HF trial. Int J Cardiol 2012 Nov 29;161(3):143-150. [doi: 10.1016/j.ijcard.2011.09.007] [Medline: 21982700]
24. Scherr D, Kastner P, Kollmann A, Hallas A, Auer J, Krappinger H, MOBITEL Investigators, et al. Effect of home-based telemonitoring using mobile phone technology on the outcome of heart failure patients after an episode of acute decompensation: randomized controlled trial. J Med Internet Res 2009 Aug 17;11(3):34-78 [FREE Full text] [doi: 10.2196/jmir.1252] [Medline: 19687005]
25. Chaudhry SI, Mattera JA, Curtis JP, Spertus JA, Herrin J, Lin Z, et al. Telemonitoring in patients with heart failure. N Engl J Med 2010 Dec 09;363(24):2301-2309. [doi: 10.1056/nejmoa1010029] [Medline: 3237394]
26. Cady R, Finkelstein S, Kelly A. A telehealth nursing intervention reduces hospitalizations in children with complex health conditions. J Telemed Telecare 2009;15(6):317-320. [doi: 10.1258/jtt.2009.090105] [Medline: 19720770]

27. Takahashi PY, Pecina JL, Upatising B, Chaudhry R, Shah ND, Van Houten H, et al. A randomized controlled trial of telemonitoring in older adults with multiple health issues to prevent hospitalizations and emergency department visits. Arch Intern Med 2012 May 28;172(10):773-779 [FREE Full text] [doi: 10.1001/archinternmed.2012.256] [Medline: 22507696]
28. Darkins A, Ryan P, Kobb R, Foster L, Edmonson E, Wakefield B, et al. Care Coordination/Home Telehealth: the systematic implementation of health informatics, home telehealth, and disease management to support the care of veteran patients with chronic conditions. Telemed J E Health 2008 Dec;14(10):1118-1126. [doi: 10.1089/tmj.2008.0021] [Medline: 19119835]
29. Noel HC, Vogel DC, Erdos JJ, Cornwall D, Levin F. Home telehealth reduces healthcare costs. Telemed J E Health 2004 Jun;10(2):170-183 [FREE Full text] [doi: 10.1089/tmj.2004.10.170] [Medline: 15319047]
30. Wakefield BJ, Holman JE, Ray A, Scherubel M, Adams MR, Hillis SL, et al. Effectiveness of home telehealth in comorbid diabetes and hypertension: a randomized, controlled trial. Telemed J E Health 2011 May;17(4):254-261. [doi: 10.1089/tmj.2010.0176] [Medline: 21476945]
31. Martín-Lesende I, Orruño E, Bilbao A, Vergara I, Cairo MC, Bayón JC, et al. Impact of telemonitoring home care patients with heart failure or chronic lung disease from primary care on healthcare resource use (the TELBIL study randomised controlled trial). BMC Health Serv Res 2013 Mar 28;13:118 [FREE Full text] [doi: 10.1186/1472-6963-13-118] [Medline: 23537332]
32. Orozco-Beltran D, Sánchez-Molla M, Sanchez JJ, Mira JJ, ValCrònic Research Group. Telemedicine in primary care for patients with chronic conditions: the valcrònic quasi-experimental study. J Med Internet Res 2017 Dec 15;19(12):e400 [FREE Full text] [doi: 10.2196/jmir.7677] [Medline: 29246881]
33. Seto E, Ware P, Logan AG, Cafazzo JA, Chapman KR, Segal P, et al. Self-management and clinical decision support for patients with complex chronic conditions through the use of smartphone-based telemonitoring: randomized controlled trial protocol. JMIR Res Protoc 2017 Nov 21;6(11):e229 [FREE Full text] [doi: 10.2196/resprot.8367] [Medline: 29162557]
34. Seto E, Leonard KJ, Cafazzo JA, Barnsley J, Masino C, Ross HJ. Developing healthcare rule-based expert systems: case study of a heart failure telemonitoring system. Int J Med Inform 2012 Aug;81(8):556-565. [doi: 10.1016/j.ijmedinf.2012.03.001] [Medline: 22465288]
35. Sultan M, Kuluski K, McIsaac WJ, Cafazzo JA, Seto E. Turning challenges into design principles: telemonitoring systems for patients with multiple chronic conditions. Health Informatics J 2019 Dec;25(4):1188-1200. [doi: 10.1177/1460458217749882] [Medline: 29320911]
36. Seto E, Ross H, Tibbles A, Wong S, Ware P, Etchells E, et al. A mobile phone-based telemonitoring program for heart failure patients after an incidence of acute decompensation (Medly-Aid): protocol for a randomized controlled trial. JMIR Res Protoc 2020 Jan 22;9(1):e15753 [FREE Full text] [doi: 10.2196/15753] [Medline: 32012116]
37. Ezekowitz JA, O'Meara E, McDonald MA, Abrams H, Chan M, Ducharme A, et al. 2017 comprehensive update of the canadian cardiovascular society guidelines for the management of heart failure. Can J Cardiol 2017 Nov;33(11):1342-1433. [doi: 10.1016/j.cjca.2017.08.022] [Medline: 29111106]
38. Ivers NM, Jiang M, Alloo J, Singer A, Ngui D, Casey CG, et al. Diabetes Canada 2018 clinical practice guidelines: key messages for family physicians caring for patients living with type 2 diabetes. Can Fam Physician 2019 Jan;65(1):14-24 [FREE Full text] [Medline: 30674509]
39. Cash J, Glass C, Gunter D, Bradley K. Hypertension (HTN), Canadian Guideline. In: Canadian Family Practice Guidelines. New York: Springer; 2019.
40. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
41. Lorig K, Stewart A, Ritter P, Gonzalez V, Laurent D, Lynch J. Outcome Measures for Health Education and Other Health Care Interventions. Thousand Oaks, CA: SAGE Publications; 1996.
42. Lorig KR, Sobel DS, Ritter PL, Laurent D, Hobbs M. Effect of a self-management program on patients with chronic disease. Eff Clin Pract 2001;4(6):256-262. [Medline: 11769298]
43. Ritter PL, Lorig K. The English and Spanish Self-Efficacy to Manage Chronic Disease Scale measures were validated using multiple studies. J Clin Epidemiol 2014 Nov;67(11):1265-1273. [doi: 10.1016/j.jclinepi.2014.06.009] [Medline: 25091546]
44. Riegel B, Lee CS, Dickson VV, Carlson B. An update on the self-care of heart failure index. J Cardiovasc Nurs 2009;24(6):485-497 [FREE Full text] [doi: 10.1097/JCN.0b013e3181b4baa0] [Medline: 19786884]
45. Bennett SJ, Oldridge NB, Eckert GJ, Embree JL, Browning S, Hou N, et al. Comparison of quality of life measures in heart failure. Nurs Res 2003;52(4):207-216. [doi: 10.1097/00006199-200307000-00001] [Medline: 12867777]
46. Venkatesh V, Morris M, Davis G, Davis F. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
47. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
48. Kraef C, van der Meirschen M, Free C. Digital telemedicine interventions for patients with multimorbidity: a systematic review and meta-analysis. BMJ Open 2020 Oct 13;10(10):e036904 [FREE Full text] [doi: 10.1136/bmjopen-2020-036904] [Medline: 33051232]

49. Ware P, Ross HJ, Cafazzo JA, Laporte A, Gordon K, Seto E. User-centered adaptation of an existing heart failure telemonitoring program to ensure sustainability and scalability: qualitative study. JMIR Cardio 2018 Dec 06;2(2):e11466 [FREE Full text] [doi: 10.2196/11466] [Medline: 31758774]
50. Olivier CB, Middleton SK, Purington N, Shashidhar S, Hereford J, Mahaffey KW, et al. Why digital health trials can fail: lessons learned from a randomized trial of health coaching and virtual cardiac rehabilitation. Cardiovas Digit Health J 2021 Apr;2(2):101-108 [FREE Full text] [doi: 10.1016/j.cvdhj.2021.01.003]
51. Gordon K, Gray CS, Dainty K, DeLacy J, Ware P, Seto E. Exploring an innovative care model and telemonitoring for the management of patients with complex chronic needs: qualitative description study. JMIR Nurs 2020;3(1):e15691 [FREE Full text] [doi: 10.2196/15691] [Medline: 34345777]
52. Gordon K, Dainty K, Gray CS, DeLacy J, Shah A, Resnick M, et al. Experiences of complex patients with telemonitoring in a nurse-led model of care: multimethod feasibility study. JMIR Nurs 2020 Sep 29;3(1):e22118 [FREE Full text] [doi: 10.2196/22118] [Medline: 34406972]
53. Kastner M, Hayden L, Wong G, Lai Y, Makarski J, Treister V, et al. Underlying mechanisms of complex interventions addressing the care of older adults with multimorbidity: a realist review. BMJ Open 2019 Apr 03;9(4):e025009 [FREE Full text] [doi: 10.1136/bmjopen-2018-025009] [Medline: 30948577]
54. de la Pisa BP, Palou-Lobato M, Calzada CM, García-Lozano MJ. Effectiveness of interventions based on telemedicine on health outcomes in patients with multimorbidity in primary care: a systematic review. Aten Primaria 2020 Dec;52(10):759-769 [FREE Full text] [doi: 10.1016/j.aprim.2019.08.004] [Medline: 31813545]
55. Bernocchi P, Vitacca M, La Rovere MT, Volterrani M, Galli T, Baratti D, et al. Home-based telerehabilitation in older patients with chronic obstructive pulmonary disease and heart failure: a randomised controlled trial. Age Ageing 2018 Dec 01;47(1):82-88. [doi: 10.1093/ageing/afx146] [Medline: 28985325]
56. Rifkin DE, Abdelmalek JA, Miracle CM, Low C, Barsotti R, Rios P, et al. Linking clinic and home: a randomized, controlled clinical effectiveness trial of real-time, wireless blood pressure monitoring for older patients with kidney disease and hypertension. Blood Press Monit 2013 Feb;18(1):8-15 [FREE Full text] [doi: 10.1097/MBP.0b013e32835d126c] [Medline: 23275313]
57. Wakefield BJ, Holman JE, Ray A, Scherubel M, Adams MR, Hills SL, et al. Outcomes of a home telehealth intervention for patients with diabetes and hypertension. Telemed J E Health 2012 Oct;18(8):575-579. [doi: 10.1089/tmj.2011.0237] [Medline: 22873700]
58. Yoo HJ, Park MS, Kim TN, Yang SJ, Cho GJ, Hwang TG, et al. A ubiquitous chronic disease care system using cellular phones and the internet. Diabet Med 2009 Jun;26(6):628-635. [doi: 10.1111/j.1464-5491.2009.02732.x] [Medline: 19538239]

Abbreviations

DM: diabetes mellitus
ED: emergency department
HCP: health care provider
HF: heart failure
HT: hypertension
MCC: multiple chronic condition
RCT: randomized controlled trial
UHN: University Health Network

Edited by R Kukafka; submitted 07.07.21; peer-reviewed by E Sadeghi-Demneh, T Greenhalgh, C Jacob; comments to author 01.10.21; revised version received 22.11.21; accepted 03.12.21; published 26.01.22.

Please cite as:
Ware P, Shah A, Ross HJ, Logan AG, Segal P, Cafazzo JA, Szacun-Shimizu K, Resnick M, Vattaparambil T, Seto E
Challenges of Telemonitoring Programs for Complex Chronic Conditions: Randomized Controlled Trial With an Embedded Qualitative Study
J Med Internet Res 2022;24(1):e31754
URL: https://www.jmir.org/2022/1/e31754
doi:10.2196/31754
PMID:35080502

Original Paper

Unit Response and Costs in Web Versus Face-To-Face Data Collection: Comparison of Two Cross-sectional Health Surveys

Elise Braekman[1], PhD; Stefaan Demarest[1], MSc; Rana Charafeddine[1], PhD; Sabine Drieskens[1], BSc; Finaba Berete[1], MPH; Lydia Gisle[1], MSc; Johan Van der Heyden[1], PhD; Guido Van Hal[2], PhD

[1]Lifestyle and chronic diseases, Epidemiology and public health, Sciensano, Brussels, Belgium

[2]Social Epidemiology and Health Policy, Antwerp University, Antwerp, Belgium

Corresponding Author:
Elise Braekman, PhD
Lifestyle and chronic diseases
Epidemiology and public health
Sciensano
Juliette Wytsmanstraat 14
Brussels
Belgium
Phone: 32 2 642 57 06
Email: elise.braekman@sciensano.be

Abstract

Background: Potential is seen in web data collection for population health surveys due to its combined cost-effectiveness, implementation ease, and increased internet penetration. Nonetheless, web modes may lead to lower and more selective unit response than traditional modes, and this may increase bias in the measured indicators.

Objective: This research assesses the unit response and costs of a web study versus face-to-face (F2F) study.

Methods: Alongside the Belgian Health Interview Survey by F2F edition 2018 (BHISF2F; net sample used: 3316), a web survey (Belgian Health Interview Survey by Web [BHISWEB]; net sample used: 1010) was organized. Sociodemographic data on invited individuals was obtained from the national register and census linkages. Unit response rates considering the different sampling probabilities of both surveys were calculated. Logistic regression analyses examined the association between mode system and sociodemographic characteristics for unit nonresponse. The costs per completed web questionnaire were compared with the costs for a completed F2F questionnaire.

Results: The unit response rate is lower in BHISWEB (18.0%) versus BHISF2F (43.1%). A lower response rate was observed for the web survey among all sociodemographic groups, but the difference was higher among people aged 65 years and older (15.4% vs 45.1%), lower educated people (10.9% vs 38.0%), people with a non-Belgian European nationality (11.4% vs 40.7%), people with a non-European nationality (7.2% vs 38.0%), people living alone (12.6% vs 40.5%), and people living in the Brussels-Capital (12.2% vs 41.8%) region. The sociodemographic characteristics associated with nonresponse are not the same in the 2 studies. Having another European (OR 1.60, 95% CI 1.20-2.13) or non-European nationality (OR 2.57, 95% CI 1.79-3.70) compared to a Belgian nationality and living in the Brussels-Capital (OR 1.72, 95% CI 1.41-2.10) or Walloon (OR 1.47, 95% CI 1.15-1.87) regions compared to the Flemish region are associated with a higher nonresponse only in the BHISWEB study. In BHISF2F, younger people (OR 1.31, 95% CI 1.11-1.54) are more likely to be nonrespondents than older people, and this was not the case in BHISWEB. In both studies, lower educated people have a higher probability of being nonrespondent, but this effect is more pronounced in BHISWEB (low vs high education level: Web, OR 2.71, 95% CI 2.21-3.39 and F2F OR 1.70, 95% CI 1.48-1.95). The BHISWEB study had a considerable advantage; the cost per completed questionnaire was almost 3 times lower (€41 [US $48]) compared with F2F data collection (€111 [US $131]).

Conclusions: The F2F unit response rate was generally higher, yet for certain groups the difference between web and F2F was more limited. Web data collection has a considerable cost advantage. It is therefore worth experimenting with adaptive mixed-mode designs to optimize financial resources without increasing selection bias (eg, only inviting sociodemographic groups who are keener to participate online for web surveys while continuing to focus on increasing F2F response rates for other groups).

(J Med Internet Res 2022;24(1):e26299) doi:10.2196/26299

KEYWORDS

health interview surveys; data collection mode; face-to-face; web; unit response; response rate; nonresponse; data collection costs; web data; health surveys; internet penetration; web survey; costs

Introduction

General population health surveys are an important data source for monitoring the health of the population and for policy making. In this regard, the Belgian Health Interview Survey (BHIS) provides periodic statistics on the health status, health care use, and health determinants of the country's population [1]. Through the BHIS, the statistics requested by the European Statistical Office (Eurostat) in the framework of the European Health Interview Survey (EHIS) are collected. Since its inception in 1997, BHIS data collection has been undertaken through face-to-face (F2F) interviews at participants' homes. In addition, participants aged 15 years and older are asked to fill out a paper-and-pencil questionnaire covering the most sensitive topics. This mix of the interview and self-administered modes is used to exploit the advantages of each [2]. F2F interviewing is particularly suited for the rather long and complex BHIS questionnaire [3,4], and the paper-and-pencil questionnaire reduces the risk of social desirability bias for sensitive topics and enhances privacy [2].

High internet penetration rates and widespread adaptation of the general public to the internet have encouraged experimentation with online HISs in developed countries to exploit the advantages of the web mode. Sending emails is the most cost-effective recruitment strategy for web surveys [5]. Even when using postal mail instead of email invitations, however, a web survey may have a considerable cost advantage over an F2F survey [6]. Moreover, web data collection shortens the duration of data collection [6,7] and is less demanding from a logistical point of view (eg, no interviewer training, no intensive interviewer follow-up) [4] compared to F2F data collection.

Unlike an F2F survey with a paper-and-pencil self-administered part, a web survey is completely self-administered, and this also provides advantages [5]. For example, the burden on respondents is expected to be lower when using the web mode since respondents can complete the questionnaire at a time convenient for them, possibly even spread over several time points, and no interviewer appointment is required [6]. While interviewers can guide and motivate respondents to complete the questionnaire optimally, their presence in itself (ie, social desirability bias), their characteristics (eg, age, gender, and ethnicity), and their interview strategies (eg, incorrect reading of the questions or inadequate probing) can affect the respondent's answers [3,8].

In spite of these advantages, the web mode is rarely used as a single mode of data collection for population-based HISs due to expected noncoverage and nonresponse issues. Although the internet access rate has increased substantially over time in Europe, noncoverage remains an issue; women, older people, and lower educated people still have lower access rates [9]. Moreover, according to a recent study internet users have a better subjective health status than internet nonusers and weighting for sociodemographic characteristics does not eliminate this observed health difference [10]. In contrast, F2F data collection is expected to be the most reliable approach to obtain a nationally representative sample [11]. In theory, any household in the country can be accessed by an interviewer, and (internet) illiterate people are not excluded by default.

Regardless of the data collection mode, HIS unit response rates have been decreasing over the past decades [12-15]. Yet several meta-analyses found that web surveys perform even worse in terms of unit response rates than surveys conducted using other modes of data collection [16-19]. The higher unit response rates in F2F surveys can be attributed to the higher perceived survey legitimacy and to the persuasiveness of having someone at the doorstep [11]. More equal participation across all sociodemographic groups is also expected when using an F2F mode than when using a web mode. Laaksonen and Heiskanen [20] reported a considerably lower unit response rate of a web compared to F2F study organized in the Finnish general population (25% vs 50%, respectively). Moreover, they showed that being older, being a nonnative Finnish speaker, and being lower educated were associated with a lower probability of responding to a web survey, whereas these sociodemographic variables were not predictors of a lower response in their F2F survey.

Such direct comparisons between F2F and web surveys based on probability samples are rare. Web data collection (using postal mail recruitment) is more frequently compared to paper-and-pencil data collection in terms of coverage and unit nonresponse rates [21-25]. The aim of this study is to compare F2F data collection with web data collection using postal mail recruitment in terms of unit response and costs of a general population health survey using a sample drawn from a national population register. An additional objective is the assessment of the sociodemographic characteristics associated with unit nonresponse using the 2 modes. The hypothesis is that in 2018, our study period, the unit response rate on the web will still be lower than the F2F unit response rate, although this difference between unit response rates will probably vary among the various sociodemographic groups in the population. It is also expected that the costs will be higher in the F2F compared to the web study.

Methods

Study Design

Alongside the traditional F2F survey (ie, the Belgian Health Interview Survey by F2F edition 2018 [BHISF2F]), a web-based survey (ie, the Belgian Health Interview Survey by Web [BHISWEB]) was organized. For both studies, authorizations were received from the Belgian privacy commission and from the ethics committee of the Ghent University Hospital.

Belgian Health Interview Survey by F2F

A cross-sectional F2F study was conducted in Belgium with a target net sample size of 11,300. The survey was organized at

household level, and households were selected based on the national population register, using a multistage clustered sampling procedure. For every selected household, 3 replacement households matched on statistical sector (ie, a subdivision of a municipality), household size, and age of the reference person were also selected. In addition to this cluster, a substitute cluster of 4 households with no matched characteristics to the first cluster was created in case of nonparticipation of all first cluster households. Sample substitution was applied during data collection: nonparticipating households were substituted, if necessary several times, by replacement households.

The households selected for the BHISF2F received a postal advance letter stating that an interviewer would visit and containing information on the BHIS. In households with a maximum of 4 members, all members were asked to participate. For households with at least 5 members, only 4 members (selected according to a systematic approach) were asked to participate.

The gross of the BHISF2F questions were administered through a computer-assisted personal interview (CAPI), but questions on the most sensitive topics were included in a paper-and-pencil self-administered questionnaire. The latter was completed by (nonproxy) respondents aged 15 years and older during the interview session. The questionnaires were available in Dutch, French, German, and English.

Data collection took place from January 2018 to January 2019 and was organized in collaboration with our fieldwork partner, the Belgian Statistical Office (Statbel). More information can be found in Multimedia Appendix 1 and in the methodological report of the BHISF2F [26].

Belgian Health Interview Survey by Web

A cross-sectional web study was conducted in Belgium with a target net sample size of 1000, a number that was feasible based on the budget set aside for this study. The national population register was used as the sampling frame. Unlike the BHISF2F, the survey was organized at the individual level and only individuals aged 16 to 85 years were selected. Moreover, individuals living in collective or institutional households and individuals living in the German-speaking region of Belgium (East Belgium, <1% of the Belgian population) were excluded from the sampling frame. A multistage clustered sampling procedure, similar to the BHISF2F, was used to select individuals. For every selected individual, 9 replacement individuals who were comparable in terms of statistical sector, sex, and age were also selected. Matched sample substitution was applied during data collection: nonrespondents were replaced, if needed several times, by replacement individuals.

Selected individuals were invited through a postal letter, and one reminder letter was sent after 7 days. The access period of the web survey was 14 days. Participants received a €10 (US $12) conditional incentive in the form of a gift voucher. The BHISWEB questionnaire was shorter than the BHISF2F questionnaire as the latter contained not only the EHIS questions but also some additional questions for national purposes. The EHIS questions corresponded with all variables requested by Eurostat in the context of EHIS wave 3 [27]. The questionnaire was available in Dutch, French, and English.

Data collection was organized together with Statbel and took place from April to November 2018, with a break during July and August. More information on the design choices made in this study can be found in Multimedia Appendix 1, Multimedia Appendix 2, and elsewhere [28].

The BHISWEB and BHISF2F studies were based on 2 mutually exclusive samples. As was made clear in the earlier descriptions, the data collection mode as well as other related design features varied between the 2 surveys. Unit response rates were therefore compared between studies using different mode systems, rather than between studies using solely different modes. To assess the differences in these unit response rates, the gross samples of the BHISF2F and the BHISWEB must be made comparable. To improve the comparability the following steps were performed: (1) only individuals invited for the BHISF2F aged 16 to 85 years and not living in an institutionalized environment or in East Belgium were included; (2) only BHISF2F individuals invited for participation between April and November 2018 (excluding the holiday months) were included to correspond with the time frame of the BHISWEB; (3) a system of weighting to adjust for the differential sample selection used in the 2 studies (gross sample for BHISF2F was n=7698 and gross sample for BHISWEB was n=6183) and consequently to adjust for the differences in their age, sex, and region distribution was applied. Weights were assigned to the people invited to the BHISWEB to make this gross sample comparable to the BHISF2F gross sample in terms of the age, sex, and region distribution. The calculation of these weights was based on cross classified data on the BHISF2F gross sample in terms of age (16-40 years, 41-65 years, and 66 years and older), sex, and region (Flemish, Brussels-Capital, and Walloon regions). The people invited to the BHISF2F all received a weight of 1.

Analyses

First, unit response rates were calculated using the weights as obtained via the method described earlier. For the BHISWEB, the unit response rate was the number of invited individuals having completed the first questions of the 3 first modules (ie, a set of questions related to the same topic) divided by the number of all invited individuals. This web response rate did not allow noncoverage due to having no internet or computer access to be disentangled from actual nonresponse. For the BHISF2F, the individual response rate and not the household response rate was calculated. More specifically, it was the number of selected individuals from the invited households who completed the first questions of the first 3 CAPI modules divided by the number of all selected individuals from the invited households. For these calculations, we did not differentiate between responses from primary selected individuals and responses from substitutes. We looked at the number of respondents in relation to the number of invited individuals (these individuals may be primary selected or substitutes). In addition to the response rate, the response rate ratio (ie, F2F unit response rate/web unit response rate) was calculated.

Demographic variables derived from the national register (sex, age group, number of household members, region of residence,

nationality, and urbanization rate) were used to calculate group-specific unit response rates for the 2 studies and response rate ratios. In addition, unit response rates and response rate ratios by level of education were calculated. Since the national register does not include information on the socioeconomic status of the invited individuals, data on the education level of invited BHISWEB and BHISF2F individuals were derived from a linkage with the Administrative Census 2011. The highest education level achieved was available in 7 categories according to the International Classification of Education and was recoded in 3 categories: low educational attainment (lower secondary education or less), intermediate educational attainment (higher secondary education and postsecondary nonhigher education), and high educational attainment (higher education). High levels of item-missingness were found for educational level (ie, this information was missing for 25.5% of all selected people). In order to get an idea of how the unit response varied with the substitution process, unit response rates per substitution wave are also presented in Table 1: wave 1 concerns the unit response rate among the initially selected individuals; waves 2 to 4 and higher concern the unit response rate among the activated substitutes in each wave.

Second, logistic regression modeling was used to study the association between unit nonresponse and mode system, sociodemographic characteristics (sex, age group, education level, number of household members, region of residence, nationality, and urbanization rate), and substitution wave (basic model).

Third, to assess whether the effect of the sociodemographic characteristics on nonresponse depended on the mode system (effect modification), interaction terms were added to the basic logistic regression model. If the interaction term was significant, we stratified the regression analyses to calculate the effect of the sociodemographic characteristics by mode system (stratified models).

Due to the high level of item-missingness on the educational level variable, regression-based multiple imputation (m=20) procedures were applied for the nonresponse analyses, presuming missingness at random. The SAS PROC MIANALYZE procedure was used for the multiple imputation. All analyses were conducted in SAS Enterprise Guide 7.1 (SAS Institute Inc). They were weighted and took into account the complex sampling designs (stratification in both studies and clustering at household level for the BHISF2F).

Last, a cost analysis was performed. The costs of the BHISWEB study were compared with the costs of an F2F study. A distinction was made between fixed costs (ie, costs regardless of the number of invitations) and variable costs (ie, costs depending on the number of invitations). The reported costs of the BHISWEB study corresponded exactly with expenditure related to BHISWEB. The costs for the F2F data collection could not be based entirely on the BHISF2F expenditure due to differences regarding the target sample size (eg, BHISF2F 11,300 vs BHISWEB 1000) and due to the specific BHISF2F financing, which included reduced tariffs from the printing company and fieldwork partner. Therefore, a cost estimation was performed for an F2F study with a target sample size of 1000 under financial conditions comparable to those in the BHISWEB study. In sum, this F2F data collection consisted of a postal advance letter and a collaboration with interviewers to contact and interview the selected individuals at home (via a CAPI including a paper-and-pencil self-administered questionnaire). Due to the higher perceived legitimacy and the persuasiveness of having someone on the doorstep, an incentive was not provided. This study was implemented with the same fieldwork partner as the BHISWEB survey. Therefore, the cost assessment for the F2F study was mainly based on their tender regarding fieldwork logistics, provision of fieldwork materials, and payment for the interviewers. Concretely, the following costs were included in the comparison: project management, information and communications technology, data warehousing, licenses for survey development, incentives, printing, packaging, postage, and interviewers.

Table 1. Weighted unit response rates of the Belgian Health Interview Survey by Web and the Belgian Health Interview Survey by F2F edition 2018.

Characteristics	BHISWEB[a] (n=6183), %	BHISF2F[b] (n=7698), %	Ratio[c] (F2F/web)
Sex			
Male	17.8	41.9	2.35
Female	18.2	44.3	2.43
Age (years)			
16-40	18.1	39.9	2.20
41-65	19.1	44.9	2.35
65+	15.4	45.1	2.93
Education			
Low	10.9	38.0	3.49
Middle	20.8	44.1	2.12
High	29.2	52.7	1.80
Nationality			
Belgian	19.7	43.7	2.22
European	11.4	40.7	3.57
Non-European	7.2	38.0	5.28
Household size			
1	12.6	40.5	3.21
2	19.3	44.3	2.30
3	17.9	42.0	2.35
≥4	20.3	44.2	2.18
Region			
Flemish	22.2	41.6	1.87
Brussels-Capital	12.2	41.8	3.43
Walloon	17.9	45.3	2.53
Urbanicity			
Urban	20.9	39.6	1.89
Suburban	18.9	46.3	2.45
Rural	21.3	53.9	2.53
Brussels-Capital	12.2	41.8	3.43
Substitution wave			
1	20.3	45.2	2.23
2	17.1	41.7	2.44
3	19.4	37.9	1.95
4	17.4	43.9	2.52
≥5	15.3	44.0	2.88

[a]BHISWEB: Belgian Health Interview Survey by Web.

[b]BHISF2F: Belgian Health Interview Survey by F2F edition 2018.

[c]Ratio: face-to-face response rate/web response rate.

Results

In total, 16.3% (1010/6183) of invited individuals participated in the BHISWEB study and 43.1% (3316/7698) of invited individuals participated in the BHISF2F study. This resulted in weighted unit response rates of 18.0% for the BHISWEB study and 43.1% for the BHISF2F study (unweighted response rates: BHISWEB 16.3%; BHISF2F 43.1%). The BHISF2F response rate was 2.39 times higher than the BHISWEB response rate. An overview of the unit response rates by sociodemographic

characteristics and substitution wave is provided in Table 1. Regardless of the sociodemographic subgroup in the population, the unit response rate was higher in the F2F study compared to the web study (ratio [F2F/web]>1). Nevertheless, the extent of this difference varied between sociodemographic subgroups. Especially for people aged 65 years and older, with low education levels, of non-Belgian nationality, living in a single-person household, or living in the Brussels-Capital region, the unit response rates were higher in the BHISF2F than in the BHISWEB study (ratio [F2F/web]>2.9). In both the BHISWEB and BHISF2F studies, unit response rates were highest among the initially selected individuals (20.3% in BHISWEB and 45.2% in BHISF2F) and not among the substitutes.

Logistic regression analysis showed that unit nonresponse was significantly higher in the web study compared to the F2F study (OR 3.53, 95% CI 3.18-3.91; Table 2: basic model). The basic model also showed that some sociodemographic characteristics (ie, education level, nationality, household size, urbanicity) were significantly associated with nonresponse.

In addition, the model including interaction terms showed both significant and nonsignificant interaction terms between mode system and sociodemographic characteristics. For sex, household size, and urbanicity, no significant interaction effects were found, which means that the association between nonresponse and these characteristics did not differ by mode system. People living with 2 (OR 0.80, 95% CI 0.70-0.90) or at least 4 other household members (OR 0.78, 95% CI 0.69-0.90) were less likely to be nonrespondents than singles, as were people living in rural (OR 0.61, 95% CI 0.49-0.77) or suburban (OR 0.78, 95% CI 0.68-0.90) areas compared to people living in urban areas (ORs based on full model).

Significant interaction terms were found for age, education level, region, and nationality. Stratified models indicating the results by mode system showed that after adjusting for all relevant covariates, there was an age effect in the F2F survey; younger people were more likely to be nonrespondents (OR 1.31, 95% CI 1.11-1.54) than older people, while no age effect was found in the web survey. The stratified analyses also showed a stronger education effect in the web study compared to the F2F study (web OR low vs high education level 2.71, 95% CI 2.21-3.39; F2F OR low vs high education level 1.70, 95% CI 1.48-1.95/web OR middle vs high education level 1.63, 95% CI 1.33-1.98; F2F OR middle vs high education level 1.44, 95% CI 1.25-1.65). Moreover, the nationality of invited people was associated with responding to a web survey but not to an F2F survey. People with another European (OR 1.60, 95% CI 1.20-2.13) or a non-European (OR 2.57, 95% CI 1.79-3.70) nationality were less likely to participate in the web survey than people with Belgian nationality. Last, a region effect was found in the web survey but not in the F2F survey. People living in the Brussels-Capital (OR 1.72, 95% CI 1.41-2.10) or Walloon (OR 1.47, 95% CI 1.15-1.87) region were less likely to participate in the web survey than people living in the Flemish Region.

Both fixed and variable costs were considerably lower for web than for F2F data collection (Table 3). The total cost per completed questionnaire was almost 3 times lower for web data collection (€41 [US $48]) compared to the F2F data collection (€111 [US $131]; Table 3). Two factors accounted for most of this cost difference: payment of the interviewers (for their completed interviews, training sessions attended, and transportation costs) and more expensive project management. The fieldwork follow-up in a web study is quite straightforward and based on automatic programming, but interviewers need extensive individual follow-up. Furthermore, a project manager is paid to deliver interviewer training and perform data checking and cleaning, which is more labor extensive in an F2F study than in a web study.

Table 2. Results of the nonresponse analyses (outcome=nonresponse), Belgian Health Interview Survey by Web, Belgian Health Interview Survey by F2F edition 2018.

	Global models		Stratified models	
	Basic[a], OR[b] (95% CI)	Full[c], OR (95% CI)	BHISWEB[d], OR (95% CI)	BHISF2F[e], OR (95% CI)
Mode system (ref[f] F2F)				
Web	3.53 (3.18-3.91)	N/A[g]	N/A	N/A
Sociodemographic characteristics				
Sex (ref male)				
Female	0.95 (0.89-1.02)	0.95 (0.89-1.02)	N/A	N/A
Age (years; ref 65+)				
16-40	1.12 (0.98-1.28)	N/A	0.86 (0.68-1.09)	1.31 (1.11-1.54)
41-65	1.02 (0.91-1.16)	N/A	0.90 (0.72-1.12)	1.11 (0.95-1.30)
Education (ref high)				
Low	2.00 (1.79-2.24)	N/A	2.71 (2.21-3.39)	1.70 (1.48-1.95)
Middle	1.51 (1.34-1.70)	N/A	1.63 (1.33-1.98)	1.44 (1.25-1.65)
Nationality (ref Belgian)				
European	1.22 (1.04-1.43)	N/A	1.60 (1.20-2.13)	1.08 (0.89-1.32)
Non-European	1.45 (1.20-1.75)	N/A	2.57 (1.79-3.70)	1.10 (0.86-1.41)
Household size (ref 1)				
2	0.80 (0.71-0.90)	0.80 (0.70-0.90)	N/A	N/A
3	0.86 (0.74-1.01)	0.87 (0.75-1.02)	N/A	N/A
≥4	0.79 (0.69-0.91)	0.78 (0.69-0.90)	N/A	N/A
Region (ref Flemish)				
Brussels-Capital	1.14 (0.99-1.31)	N/A	1.72 (1.41-2.10)	0.87 (0.72-1.05)
Walloon	1.09 (0.96-1.25)	N/A	1.47 (1.15-1.87)	0.93 (0.79-1.09)
Urbanicity (ref urban)				
Rural	0.65 (0.52-0.81)	0.61 (0.49-0.77)	N/A	N/A
Suburban	0.85 (0.75-0.97)	0.78 (0.68-0.90)	N/A	N/A
Substitution wave (ref 1)				
2	1.16 (1.02-1.33)	1.16 (1.02-1.32)	N/A	N/A
3	1.20 (1.03-1.40)	1.19 (1.02-1.38)	N/A	N/A
4	1.10 (0.93-1.31)	1.09 (0.92-1.30)	N/A	N/A
≥5	1.15 (0.98-1.36)	1.04 (0.88-1.24)	N/A	N/A

[a]Basic model: ORs based on logistic regression model with nonresponse as outcome and mode system and sociodemographic characteristics and substitution wave as independent variables.

[b]OR: odds ratio.

[c]Full model: ORs based on logistic regression model with nonresponse as outcome and mode system and sociodemographic characteristics, substitution wave, and significant interaction terms between mode system and sociodemographic characteristics as independent variables.

[d]BHISWEB (Belgian Health Interview Survey by Web): ORs based on logistic regression model with nonresponse as outcome and sociodemographic characteristics as independent variable for the BHISWEB study to show the stratified results for sociodemographic characteristics with significant interaction terms in the full model.

[e]BHISF2F (Belgian Health Interview Survey by F2F edition 2018): ORs based on logistic regression model with nonresponse as outcome and sociodemographic characteristics as independent variable for the BHISF2F study to show the stratified results for sociodemographic characteristics with significant interaction terms in the full model.

[f]ref: reference.

[g]N/A: not applicable.

Table 3. Cost figures[a] for web versus F2F data collection with a target sample size of 1000, Belgian Health Interview Survey by Web and F2F mode system.

	BHISWEB[b] (n=1010), €	F2F[c] mode system (n=1000), €
Fixed costs		
Fieldwork logistics		
Project management[d]	1641	35,040
ICT[e]	3282	3282
Data ware housing	4922	4922
Sampling	1641	1641
Licenses[f]		
CAPI[g] software	N/A[h]	5375
Web survey software	8740	N/A
Total fixed costs	20,226	50,260
Fixed costs per completed questionnaire	20	50
Variable costs		
Incentive payments	10,212	N/A
Printing, packaging and postage		
Invitation letter (+folder)	5387	N/A
Reminder letter	4829	N/A
Incentive letter	880	N/A
Interviewer materials and advance letter (+folder)	N/A	4391
Interviewer laptops	N/A	5200
Interviewer payments	N/A	51,057
Total variable costs	21,308	60,648
Variable costs per completed questionnaire	21	61
Total costs		
Total fixed and variable costs	41,534	110,908
Costs per completed questionnaire	41	111

[a]Costs do not include salaries for researchers.

[b]BHISWEB: Belgian Health Interview Survey by Web.

[c]F2F: face-to-face.

[d]Project management includes costs for data control, follow-up, and training of interviewers; testing the programs; and salaries of project managers and administrative employees.

[e]Information and communication technology (ICT) includes the costs for developing an ICT infrastructure to organize the fieldwork.

[f]Costs associated with the training courses for developing the computer-assisted personal interview (CAPI) and web questionnaires were not considered since these development skills had already been acquired.

[g]CAPI: computer-assisted personal interview.

[h]N/A: not applicable.

Discussion

Principal Findings

In the context of the BHIS, web and F2F data collection were compared in terms of unit response, taking into account the different sociodemographic groups in the population and financial costs.

A response rate of 18.0% was obtained in the BHISWEB study and 43.1% in the BHISF2F study, making the web survey response almost 2.5 times lower compared to the F2F survey response. For all sociodemographic subgroups in the population, the unit response rate was higher in the F2F study compared to the web study. Nevertheless, the difference between web and F2F was more pronounced among people aged 65 years and older, with low education level, of non-Belgian nationality, living in a single-person household, and living in the

Brussels-Capital region. When taking into account only the individuals initially invited (not the substitutes of nonrespondents), the response rates were higher (20.3% in BHISWEB and 45.2% in BHISF2F). This can be explained by the fact that substitutes of hard-to-reach individuals are selected because they have similar sociodemographic characteristics and, consequently, they also have a higher chance on nonresponse [29].

Different sociodemographic characteristics were associated with nonresponding (or, conversely, with responding) in the BHISWEB compared to the BHISF2F survey. Having a non-Belgian nationality and living in the Brussels-Capital or Walloon regions were associated with a higher nonresponse rate in our web survey but not in our F2F survey. Age, on the other hand, was associated with nonresponse in the F2F study but not in the web study. In the F2F study, older people were less likely to be nonrespondents. In both studies, people with low or intermediate educational levels were less likely to respond than people with high educational levels, but this effect was stronger in the web study. The association between household size and urbanicity and nonresponse did not differ between the studies. Singles and people living in urban areas were less likely to respond to both studies.

The BHISWEB study has a considerable cost advantage over the F2F study; the total cost per completed questionnaire was almost 3 times lower (€41 [US $48]) compared to the F2F data collection (€111 [US $131]).

Strengths and Limitations

A positive aspect of this study is the comparison of an F2F study and a web study using 2 random samples drawn from the national population register and the large amount of sociodemographic information available from both participating and nonparticipating sample members. The latter is not only due to the information obtained through the national register but also due to the efforts made by linking to the Administrative Census 2011 to obtain educational information. The use of this data is not perfect since there is a time delay of 7 years between our data collection and the last administrative census and it includes a considerable number of missing values regarding the highest level of educational attainment achieved. In order to address this item-missingness, multiple imputation procedures were applied under the missingness at random assumption.

One limitation is the fact that noncoverage due to having no internet or computer access could not be disentangled from actual nonresponse in our web survey. In Belgium, 87% of households with at least one household member aged between 16 and 74 years had access to the internet at the time of the study [30], so part of the unit nonresponse is in fact linked to noncoverage. A second limitation was the strict focus on unit response rates (and unit response rate differences between sociodemographic subgroups), although there is evidence that low response rates do not necessarily lead to large nonresponse bias (ie, the difference between the expected estimate based on the respondents and the true value in the population) [31]. A Danish interview-administered study also found that although the nonresponse rate was higher among people with low socioeconomic status, no significant association was found between health status and nonresponse [32]. Proxy measures for health were used in this study: register data on hospital admission costs and dispensed prescription medicine costs. Moreover, increasing fieldwork efforts might increase the response rate, but this is not necessarily a cost-effective way of minimizing survey error [33]. Next to unit response, assessing factors related to questionnaire breakoff and item response would be of interest, as these are less commonly studied [34].

Comparison With Prior Work

The response rate difference of 26 percentage points between F2F versus web was higher than the mean difference of 12 percentage points reported in the recent meta-analyses of Daikeler et al [19], who compared response rates between the web and other survey modes. Nonetheless, our web survey was organized in the general population among newly recruited individuals (no panel members), and postal mail instead of email invitations were used. These factors are known to contribute to a higher response rate difference between web surveys and surveys organized using other modes [19]. For all sociodemographic groups, a higher response was found in the F2F study versus web study, which indicates that F2F data collection is still the most appropriate way to achieve acceptable response rates among all sociodemographic groups [7].

In addition, we found that the association between sociodemographic characteristics and nonresponse varied between the BHISWEB and BHISF2F surveys. In line with the results of other studies [6,20], we found, for example, that people of non-Belgian origin participate less in web surveys. This can be explained by their lower internet access rates [35-37]. Moreover, a web questionnaire is self-administered, which means that respondents should be not only internet literate but also capable of reading and fully understanding questions in the official national language. This is not the case when using F2F data collection; respondents must only be able to understand the questions posed by the interviewers, but these interviewers can clarify and repeat questions when needed. Moreover, interviewers can motivate people who are less fluent in the official national language to participate by highlighting the importance of the study in simple language and by referring to the help they will offer during the interview process. This can explain why our results showed that nationality was not associated with a lower F2F response.

Second, young people had higher nonresponse rates in our F2F survey than those in older age groups. This age effect was not found in the web study. Additional analysis showed that the higher F2F nonresponse was related only to their higher noncontact rates because refusal rates did not differ between different age groups. Previous studies also found higher noncontact rates among younger age groups [38,39]. This is attributed to the fact that this working-age population group is less likely to be at home when an interviewer contacts them than people older than 65 years, who are most often retired [38]. In a web survey, these interviewer contacts at home are not required, which may explain why no age effect was found in the web survey. Moreover, younger people have a high probability of meeting the necessary conditions (eg, internet

access) and having the skills to participate in a web survey [5,9,35,36,40].

Third, people with low education levels are less likely to participate in surveys, regardless of the data collection modes used [6,13,20,23,41-43]. Our unit nonresponse analyses by mode system also confirmed this since lower educated people participated less in both studies. Nevertheless, this socioeconomic difference was stronger in the web survey than in the F2F survey. Reasons for this could be lower internet access rates [9,36,37] and less frequent internet use among low educated people [40]. Moreover, this socioeconomic group may have a greater need for interviewers to explain the importance of the survey and to motivate them to participate.

A considerable cost advantage of web versus F2F data collection was reported in this paper. Most other cost comparison studies have also found a major cost advantage when using the web versus other modes (F2F, telephone, paper). A cost comparison conducted in the framework of a cross-sectional parental survey on the mental health of children showed that web survey data collection (using postal mail recruitment and including one reminder letter) was 4 times cheaper than F2F data collection [6]. Substantial cost advantages (half of the cost) were also reported for web data collection (using mail invitations) compared to paper-and-pencil data collection in the context of a parental survey on children's health status [22]. Sinclair et al [21] reported that a web survey (using postal mail invitations) offered a considerable cost advantage compared to a telephone survey organized in the context of a community-based health survey. Nevertheless, their paper-and-pencil survey using mail recruitment had lower costs than their web survey due to the high costs associated with web survey development. By using email instead of postal mail invitations, the cost advantage of a web survey over other modes would even be more pronounced. In most countries, emails cannot be sent to a random sample drawn from the general population because email addresses are not available for researchers to contact potential respondents. Denmark is an exception since a large proportion of Danish citizens have a mandatory digital mailbox that is used for communication with public authorities and to request survey participation [42].

Conclusion

The use of F2F data collection should be preferred over the use of single-mode web data collection for population-based HISs. This recommendation is based on (1) the considerably lower unit response rates of the web survey compared to the F2F survey (18% vs 43%), (2) markedly lower response rates from some specific sociodemographic groups in the web survey versus F2F survey (ie, older people, low educated people, people of foreign nationalities, and people living alone), and (3) the nonresponse analyses which showed that certain sociodemographic groups—people with low education level, of non-Belgian nationalities, and living in the Brussels-Capital or Walloon regions—were more disadvantaged in a web study compared to an F2F study.

Lower response rates can induce more bias in measured HIS indicators because there is a greater chance that web respondents and nonrespondents show differences not only in terms of these sociodemographic characteristics but also in terms of their health status and health behavior characteristics. Moreover, if response rates show greater differences between various sociodemographic groups, this will affect comparisons between different sociodemographic groups. Single-mode web data collection would therefore better be restricted to specific target groups with universal internet access (eg, university students, online panel members) and should preferably not be used for surveys organized in the general population. This study did, however, show that web data collection offered a considerable cost advantage compared to F2F data collection.

Recommendations and Future Prospects

In order to benefit from this cost advantage to some extent without increasing the risk of nonresponse bias, the web mode could be integrated in a mixed-mode design. This approach, in which some respondents complete the questionnaire on the web and other respondents (ie, those unwilling/incapable to participate online) use another mode, is already being tested and used in multiple European HISs [44-46]. A specific mixed-mode methodology, push-to-web, could, for example, be considered: people would first be invited by postal mail to participate online and they would then be contacted by an interviewer only in case of nonparticipation [47].

Based on the results of this study, experimenting with adaptive survey designs in which different sample members are assigned to different data collection modes can also be recommended. A potential strategy could be to invite only sociodemographic groups more eager to participate online for a web HIS while continuing to focus on increasing F2F response rates for the other sociodemographic groups. Tailoring the HIS data collection to different sociodemographic groups could reduce the nonresponse bias without increasing the costs.

When considering mode changes, it might also be useful to experiment with adapted recruitment procedures. Elements to take into account in future research could include the incentives for participation, the number and type of reminders, the use of tailored invitation letters based on age group, and the allocation of experienced interviewers to work with difficult to reach subpopulations.

When evaluating new designs, it might be worthwhile to focus not only on unit response rates but also on other less ambiguous indicators of nonresponse bias. These could include the calculation of R-indicators (R stands for representativeness), indicators that measure the similarity between the respondents to a survey and the sample or the population under investigation [48].

Acknowledgments

The study was funded by the internal budget of Sciensano. The authors wish to thank the Belgian Statistical Office for organizing the fieldwork.

Authors' Contributions

EB conducted the literature searches and summaries of previous related work, undertook the statistical analyses, interpreted the results, wrote the initial version of the manuscript, and conducted the revisions. S Drieskens developed the web and computer-assisted personal interview questionnaires. EB, S Demarest, and RC conceptualized the study. JVdH and FB contributed to the methods used for the statistical analyses. S Demarest, RC, and GVH were responsible for supervision. S Demarest, RC, S Drieskens, FB, LG, JVdH, and GVH substantively revised the earlier drafts of the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Sampling, sample substitution, recruitment procedure, and proxy interviewing.
[DOCX File , 65 KB - jmir_v24i1e26299_app1.docx]

Multimedia Appendix 2
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[DOCX File , 19 KB - jmir_v24i1e26299_app2.docx]

References

1. Demarest S, Van der Heyden J, Charafeddine R, Drieskens S, Gisle L, Tafforeau J. Methodological basics and evolution of the Belgian health interview survey 1997-2008. Arch Public Health 2013;71(1):24 [FREE Full text] [doi: 10.1186/0778-7367-71-24] [Medline: 24047278]
2. De Leeuw D. To mix or not to mix data collection modes in surveys. J Off Stat 2005;21(2):233-255 [FREE Full text]
3. Loosveldt G. Face-to-face interviews. In: De Leeuw ED, Hox JJ, Dillman DA, editors. International Handbook of Survey Methodology. New York: Taylor & Francis Group; 2008:210-220.
4. De Leeuw ED. Choosing the method of data collection. In: De Leeuw ED, Hox JJ, Dillman DA, editors. International Handbook of Survey Methodology. New York: Taylor & Francis Group; 2008:113-135.
5. Callegaro M, Manfreda K, Vehovar V. Web Survey Methodology. London: Sage; 2015.
6. Heiervang E, Goodman R. Advantages and limitations of web-based surveys: evidence from a child mental health survey. Soc Psychiatry Psychiatr Epidemiol 2011;46(1):69-76. [doi: 10.1007/s00127-009-0171-9] [Medline: 19921078]
7. Blair J, Czaja R, Blair E. Designing Surveys: A Guide to Decisions and Procedures. 3rd edition. Thousand Oaks: SAGE Publications; 2013.
8. Tourangeau R, Rips L, Rasinski K. The Psychology of Survey Response. Cambridge: Cambridge University Press; 2000.
9. Mohorko A, De Leeuw E, Hox J. Internet coverage and coverage bias in Europe: developments across countries and over time. J Off Stat 2013;29(4):609-622. [doi: 10.2478/jos-2013-0042]
10. Schnell R, Noack M, Torregroza S. Differences in general health of internet users and non-users and implications for the use of web surveys. Surv Res Methods 2017;11(2):105-123. [doi: 10.18148/srm/2017.v11i2.6803]
11. Roberts C. Mixing modes of data collection in surveys: a methodological review. 2007. URL: http://eprints.ncrm.ac.uk/418/1/MethodsReviewPaperNCRM-008.pdf [accessed 2021-12-10]
12. Aromaa A, Koponen P, Tafforeau J, Vermeire C, HIS/HES Core Group. Evaluation of health interview surveys and health examination surveys in the European Union. Eur J Public Health 2003;13(3 Suppl):67-72. [doi: 10.1093/eurpub/13.suppl_1.67] [Medline: 14533752]
13. Galea S, Tracy M. Participation rates in epidemiologic studies. Ann Epidemiol 2007;17(9):643-653. [doi: 10.1016/j.annepidem.2007.03.013] [Medline: 17553702]
14. Morton SMB, Bandara DK, Robinson EM, Carr PEA. In the 21st Century, what is an acceptable response rate? Aust N Z J Public Health 2012;36(2):106-108. [doi: 10.1111/j.1753-6405.2012.00854.x] [Medline: 22487341]
15. Tolonen H, Helakorpi S, Talala K, Helasoja V, Martelin T, Prättälä R. 25-year trends and socio-demographic differences in response rates: Finnish adult health behaviour survey. Eur J Epidemiol 2006;21(6):409-415. [doi: 10.1007/s10654-006-9019-8] [Medline: 16804763]
16. Manfreda KL, Bosnjak M, Berzelak J, Haas I, Vehovar V. Web surveys versus other survey modes: a meta-analysis comparing response rates. Int J Market Res 2008;50(1):79-104. [doi: 10.1177/147078530805000107]

17. Shih T, Fan X. Response rates and mode preferences in web-mail mixed-mode surveys: a meta-analysis. Int J Internet Sci 2007;2(1):59-82 [FREE Full text]
18. Blumenberg C, Barros AJD. Response rate differences between web and alternative data collection methods for public health research: a systematic review of the literature. Int J Public Health 2018;63(6):765-773. [doi: 10.1007/s00038-018-1108-4] [Medline: 29691594]
19. Daikeler J, Bosnjak M, Lozar MK. Web versus other survey modes: an updated and extended meta-analysis comparing response rates. J Surv Stat Methodol 2020;8(3):513-519. [doi: 10.1093/jssam/smz008]
20. Laaksonen S, Heiskanen M. Comparison of three modes for a crime victimization survey. J Survey Stat Methodol 2014;2(4):459-483. [doi: 10.1093/jssam/smu018]
21. Sinclair M, O'Toole J, Malawaraarachchi M, Leder K. Comparison of response rates and cost-effectiveness for a community-based survey: postal, internet and telephone modes with generic or personalised recruitment approaches. BMC Med Res Methodol 2012;12:132 [FREE Full text] [doi: 10.1186/1471-2288-12-132] [Medline: 22938205]
22. Hohwü L, Lyshol H, Gissler M, Jonsson SH, Petzold M, Obel C. Web-based versus traditional paper questionnaires: a mixed-mode survey with a Nordic perspective. J Med Internet Res 2013;15(8):e173 [FREE Full text] [doi: 10.2196/jmir.2595] [Medline: 23978658]
23. Dannetun E, Tegnell A, Giesecke J. Parents' attitudes towards hepatitis B vaccination for their children. A survey comparing paper and web questionnaires, Sweden 2005. BMC Public Health 2007;7:86 [FREE Full text] [doi: 10.1186/1471-2458-7-86] [Medline: 17511891]
24. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of Internet versus paper-and-pencil versions. J Med Internet Res 2007;9(3):e25 [FREE Full text] [doi: 10.2196/jmir.9.3.e25] [Medline: 17942387]
25. Källmén H, Sinadinovic K, Berman AH, Wennberg P. Risky drinking of alcohol in Sweden: a randomized population survey comparing web- and paper-based self-reports. Nordic Stud Alcohol Drugs 2017;28(2):123-130. [doi: 10.2478/v10199-011-0013-4]
26. Demarest S, Berete F, Charafeddine R, Van der Heyden J. Gezondheidsenquête 2018: Methodologie. Brussel: Sciensano; 2019. URL: https://his.wiv-isp.be/nl/Gedeelde%20%20documenten/Method_NL_2018.pdf [accessed 2022-01-06]
27. Eurostat. European Health Interview Survey (EHIS wave 3) Methodological Manual. Luxemburg: Publication Office of the European Union; 2018. URL: https://ec.europa.eu/eurostat/documents/3859598/8762193/KS-02-18-240-EN-N.pdf/5fa53ed4-4367-41c4-b3f5-260ced9ff2f6 [accessed 2021-12-10]
28. Braekman E, Charafeddine R, Demarest S, Drieskens S, Tafforeau J, Van der Heyden J, et al. Is the European Health Interview Survey online yet? Response and net sample composition of a web-based data collection. Eur J Public Health 2020;30(3):567-573. [doi: 10.1093/eurpub/ckz206] [Medline: 31697353]
29. Demarest S, Molenberghs G, Van der Heyden J, Gisle L, Van Oyen H, de Waleffe S, et al. Sample substitution can be an acceptable data-collection strategy: the case of the Belgian Health Interview Survey. Int J Public Health 2017;62(8):949-957. [doi: 10.1007/s00038-017-0976-3] [Medline: 28470413]
30. Beschikbaarheid van internet in het huishouden per gewest 2006-2018. Statbel. 2019. URL: https://bestat.statbel.fgov.be/bestat/crosstable.xhtml?view=d136ea24-9d50-421e-871a-7b807d163236 [accessed 2021-12-10]
31. Peytcheva E, Groves R. Using variation in response rates of demographic subgroups as evidence of nonresponse bias in survey estimates. J Off Stat 2009;25(2):193-201.
32. Ekholm O, Gundgaard J, Rasmussen NKR, Hansen EH. The effect of health, socio-economic position, and mode of data collection on non-response in health interview surveys. Scand J Public Health 2010;38(7):699-706. [doi: 10.1177/1403494810382474] [Medline: 20851845]
33. Sturgis P, Williams J, Brunton-Smith I, Moore J. Fieldwork effort, response rate, and the distribution of survey outcomes: a multilevel meta-analysis. Public Opin Q 2017;81(2):523-542. [doi: 10.1093/poq/nfw055]
34. Blumenberg C, Zugna D, Popovic M, Pizzi C, Barros AJD, Richiardi L. Questionnaire breakoff and item nonresponse in web-based questionnaires: multilevel analysis of person-level and item design factors in a birth cohort. J Med Internet Res 2018;20(12):e11046 [FREE Full text] [doi: 10.2196/11046] [Medline: 30530454]
35. Leenheer J, Scherpenzeel A. Does it pay off to include non-internet households in an internet panel? Int J Internet Sci 2013;8(1):17-29 [FREE Full text]
36. Sterrett D, Malato D, Benz J, Tompson T, English N. Assessing changes in coverage bias of web surveys in the United States. Public Opin Q 2017;81(S1):338-356. [doi: 10.1093/poq/nfx002]
37. Couper MP, Kapteyn A, Schonlau M, Winter J. Noncoverage and nonresponse in an Internet survey. Soc Sci Res 2007;36(1):131-148. [doi: 10.1016/j.ssresearch.2005.10.002]
38. Durrant G, Steele F. Multilevel modelling of refusal and non-contact in household surveys: evidence from six UK Government surveys. J R Stat Soc Ser A Stat Soc 2009;172(2):361-381. [doi: 10.1111/j.1467-985x.2008.00565.x]
39. Van der Heyden J, Demarest S, Van Herck K, De Bacquer D, Tafforeau J, Van Oyen H. Association between variables used in the field substitution and post-stratification adjustment in the Belgian health interview survey and non-response. Int J Public Health 2014;59(1):197-206. [doi: 10.1007/s00038-013-0460-7] [Medline: 23619721]

40. ICT-gebruik bij individuen. Statbel. URL: https://statbel.fgov.be/nl/themas/huishoudens/ict-gebruik-huishoudens#figures [accessed 2021-12-10]
41. Gajic A, Cameron D, Hurley J. The cost-effectiveness of cash versus lottery incentives for a web-based, stated-preference community survey. Eur J Health Econ 2012;13(6):789-799. [doi: 10.1007/s10198-011-0332-0] [Medline: 21691841]
42. Ebert JF, Huibers L, Christensen B, Christensen MB. Paper- or web-based questionnaire invitations as a method for data collection: cross-sectional comparative study of differences in response rate, completeness of data, and financial cost. J Med Internet Res 2018 Jan 23;20(1):e24 [FREE Full text] [doi: 10.2196/jmir.8353] [Medline: 29362206]
43. Demarest S, Van der Heyden J, Charafeddine R, Tafforeau J, Van Oyen H, Van Hal G. Socio-economic differences in participation of households in a Belgian national health survey. Eur J Public Health 2013;23(6):981-985. [doi: 10.1093/eurpub/cks158] [Medline: 23183496]
44. Jensen HAR, Ekholm O, Davidsen M, Christensen AI. The Danish health and morbidity surveys: study design and participant characteristics. BMC Med Res Methodol 2019;19(1):91 [FREE Full text] [doi: 10.1186/s12874-019-0733-9] [Medline: 31053088]
45. Lange C, Finger JD, Allen J, Born S, Hoebel J, Kuhnert R, et al. Implementation of the European health interview survey (EHIS) into the German health update (GEDA). Arch Public Health 2017;75:40 [FREE Full text] [doi: 10.1186/s13690-017-0208-6] [Medline: 28936356]
46. Mauz E, Hoffmann R, Houben R, Krause L, Kamtsiuris P, Gößwald A. Mode equivalence of health indicators between data collection modes and mixed-mode survey designs in population-based health interview surveys for children and adolescents: methodological study. J Med Internet Res 2018;20(3):e64 [FREE Full text] [doi: 10.2196/jmir.7802] [Medline: 29506967]
47. Lynn P. Evaluating push-to-web methodology for mixed-mode surveys using address-based samples. Surv Res Methods 2020 2020;14(1):19-30. [doi: 10.18148/SRM/2020.V14I1.7591]
48. Schouten B, Cobben F, Bethlehem J. Indicators for the representativeness of survey response. Surv Methodol 2009;35(1):110-113.

Abbreviations

BHIS: Belgian Health Interview Survey
BHISF2F: Belgian Health Interview Survey by F2F edition 2018
BHISWEB: Belgian Health Interview Survey by Web
CAPI: computer-assisted personal interview
EHIS: European Health Interview Survey
Eurostat: European Statistical Office
F2F: face-to-face
OR: odds ratio
Statbel: Belgian Statistical Office

Edited by R Kukafka, G Eysenbach; submitted 06.12.20; peer-reviewed by M Callegaro, M Jordan-Marsh; comments to author 16.03.21; revised version received 30.03.21; accepted 05.10.21; published 07.01.22.

Please cite as:
Braekman E, Demarest S, Charafeddine R, Drieskens S, Berete F, Gisle L, Van der Heyden J, Van Hal G
Unit Response and Costs in Web Versus Face-To-Face Data Collection: Comparison of Two Cross-sectional Health Surveys
J Med Internet Res 2022;24(1):e26299
URL: https://www.jmir.org/2022/1/e26299
doi:10.2196/26299
PMID:34994701

Original Paper

The Contribution of the Internet to Reducing Social Isolation in Individuals Aged 50 Years and Older: Quantitative Study of Data From the Survey of Health, Ageing and Retirement in Europe

Patrícia Silva[1*], PhD; Alice Delerue Matos[2*], PhD; Roberto Martinez-Pecino[3*], PhD

[1]Communication and Society Research Centre, University of Minho, Braga, Portugal

[2]Department of Sociology, Institute of Social Sciences, University of Minho, Braga, Portugal

[3]Departamento de Psicología Social, Facultad de Comunicación, Universidad Sevilla, Seville, Spain

[*]all authors contributed equally

Corresponding Author:
Patrícia Silva, PhD
Communication and Society Research Centre
University of Minho
Campus Gualtar, Social Sciences Institute
Braga, 4710-057
Portugal
Phone: 351 253604212
Email: patriciasilva@ics.uminho.pt

Abstract

Background: Social isolation has a negative impact on the quality of life of older people; therefore, studies have focused on identifying its sociodemographic, economic, and health determinants. In view of the growing importance of the internet as a means of communication, it is essential to assess whether internet use interferes with social isolation.

Objective: This study specifically aims to clarify the relationship between internet use and social isolation of individuals aged ≥50 years, for which other surveys present contradictory results.

Methods: We performed logistic regression analysis with social isolation as the dependent variable, internet use as the interest variable, and several other sociodemographic, economic, and health characteristics of the individuals as control variables. The sample size was 67,173 individuals aged 50 years and older from 17 European countries (Portugal, Greece, Italy, Spain, Denmark, Sweden, Austria, Belgium, France, Germany, Switzerland, Luxemburg, Poland, Czech Republic, Slovenia, Estonia, and Croatia) plus Israel, who were interviewed in the Survey of Health, Ageing and Retirement in Europe (SHARE), wave 6.

Results: The results show that countries differ in the level of social isolation and rate of internet use by individuals aged 50 years and older. They also evidence that in most of the countries analyzed, social isolation of internet users was lower compared to that of nonusers after controlling for a set of sociodemographic, economic, and health characteristics of the individuals that have been previously described in the literature as determinants of social isolation. Indeed, on average, although 31.4% of individuals in the nonuser group experienced high social isolation, only 12.9% of individuals who used the internet experienced this condition.

Conclusions: Internet users show lower social isolation. This result underlines the importance of promoting e-inclusion in Europe as a way to counter social isolation of individuals aged 50 years and older.

(J Med Internet Res 2022;24(1):e20466) doi:10.2196/20466

KEYWORDS

social isolation; internet; 50+ individuals; e-inclusion; SHARE

Introduction

Social isolation has been defined as the objective situation of individuals with small social networks and reduced frequency of contact who do not take part in social activities [1-4]. According to the literature, social isolation is associated with an increased risk of mortality [2,5-7]. In terms of health, it has been associated with a greater risk of developing chronic

diseases [8] and cardiovascular disease [3,7], as well as increased risk of physical inactivity, tobacco consumption, and various other risk behaviors [3,9]. Furthermore, in old age, social isolation is associated with increased feelings of loneliness [3].

According to the literature, social isolation tends to increase as individuals age [3,10,11]. Indeed, some events that are more frequent at an older age, such as retirement, the development or worsening of health and mobility limitations, or a change in residence of the individuals with whom they socialize [2,12-16], tend to affect older people's ability to maintain their social networks and may also cause cognitive decline [17] and increase social isolation [2].

Social isolation is determined by a set of sociodemographic, economic, and health characteristics of individuals [3,10,13,18,19]. In addition to these determinants, the role of the internet in reducing social isolation is increasingly being discussed [20] because, as an important means of communication, this technology can facilitate interpersonal contact at a stage of life when social networks tend to be restructured [12,21-24]. There are often obstacles to the use of the internet by people of advanced age [25-31]. However, internet use by older adults has been associated with feelings of well-being and social support [32-35], as well as with improved quality of life [36]. Nevertheless, the relationship between internet use and social isolation is unclear, and there is open debate in the literature. On the one hand, some studies conclude that internet use is associated with a decrease in social isolation [11,20,37,38]. In this regard, they claim that the internet has been conducive to successful interactions [39-41] and contributes to maintaining social ties [41-44] and increasing the frequency of contact [45], as well as optimizing the effect of social networks on the quality of life of older people [36]. On the other hand, other studies conclude that using this technology does not reduce social isolation [46]. Furthermore, in some specific cases, its use is actually associated with a greater risk of social isolation. In this regard, problematic and addictive uses of the internet may be related to social isolation [47]. Similarly, some studies support the "time displacement" thesis, which states that the longer individuals surf the internet, the less time they interact with family and friends, as the time spent on one activity cannot be spent on another [48].

The inconsistent results in the literature underscore the need for more research into the relationship between internet use and social isolation [20,49], in which this concept is clearly defined [50] and which draws on large samples [38,51]. This study specifically aims to contribute to this goal by focusing on the relationship between internet use and social isolation after controlling for sociodemographic, economic, and health variables and selecting a target population of individuals aged 50 years or older residing in 17 European countries and Israel.

Methods

Data and Sample

This study focuses on 67,173 individuals aged 50 years and older who were interviewed in the Survey of Health, Ageing and Retirement in Europe (SHARE) Project (wave 6) in Austria (n=3358), Germany (n=4347), Sweden (n=3881), Spain (n=5560), Italy (n=5211), France (n=3870), Denmark (n=3661), Greece (n=4814), Switzerland (n=2772), Belgium (n=5700), Israel (n=2013), Czech Republic (n=4793), Poland (n=1802), Luxembourg (n=1543), Portugal (n=1662), Slovenia (n=4186), Estonia (n=5557), and Croatia (n=2443). Details of the SHARE study in Europe have been described elsewhere [52]. Briefly, in wave 6 (2015), a survey of a representative sample of the noninstitutionalized population aged 50 years or older was conducted. Interviews were face-to-face and took place in the household. Trained interviewers conducted the interviews using a computer-assisted personal interviewing program.

The SHARE project, coordinated internationally by the Max Planck Institute for Social Law and Social Policy (Germany), has been approved by the Ethics Council of the Max Planck Society for the Advancement of Science.

Measures

Dependent Variable: Social Isolation

As in other research [3,4,53], social isolation was computed using a 5-item index. Individuals scored 1 point if they did not live with a partner; 1 point if they did not belong to any organizations, clubs, or religious groups; and 1 point for having less than monthly contact with friends, family, or children. Scores on the index ranged from 0 to 5, with higher scores indicating a greater degree of isolation. Adopting the criteria of previous studies [2,54-56], we dichotomized the social isolation index to distinguish between low (score <2) and high (score ≥2) levels of social isolation.

Low social isolation was coded as 0, and high social isolation was coded as 1.

Independent Variable: Internet Use

This is a dichotomous variable relating to use of the internet at least once in the last 7 days to send and receive emails, to search for information, to shop, or for any other purpose. This variable distinguishes individuals who use the internet (1) from individuals who do not use it (0).

Covariables: Sociodemographic and Economic Variables

The following covariables were considered in this research:

- Age; gender: female (1) and male (0); years of schooling; self-perception of financial stress: "great difficulty" or "some difficulty" in coping with monthly expenses (0), "easy" or "very easy" to cover monthly expenses (1).
- Geographical distance from social network: scores ranged from "in the same household" (1) to "more than 500 km away" (8).
- Loneliness: assessed through a short version of the Revised UCLA Loneliness Scale (R-UCLA) [57]. The scale includes 3 questions: "How much of the time do you feel you lack companionship?" "How much of the time do you feel left out?" and "How much of the time do you feel isolated from others?" The answer options range from 1 (hardly ever) to 3 (often). The 3 items form a scale that ranges from 3 to 9

points, in which high values represent higher levels of loneliness.

We also considered health variables:

- Depressive symptoms, evaluated by the EURO-D scale [58]. The EURO-D scale ranges from 0 to 12 points, which refer to the presence or absence of 12 symptoms of depression, such as suicidal thoughts [58]; as in previous studies [59], we distinguished between more symptoms (1) and individuals with lower scores (0). A score of 4 or more symptoms is considered as a cutoff point to identify major depression [60].
- Limitations in activities of daily living [61,62], which refer to the presence-absence of difficulties performing on one's own any of 6 daily living activities, such as bathing, dressing, and toileting; we distinguished between individuals who reported 1 or more limitations (1) and individuals who reported no limitations (0).
- Physical inactivity: Two items were used to assess the level of physical activity. The first item assessed vigorous physical activity ("We would like to know about the type and amount of physical activity you do in your daily life. How often do you engage in vigorous physical activity, such as sports, heavy housework, or a job that involves physical labor?"). The second item assessed moderate physical activity ("How often do you engage in activities that require a low or moderate level of energy such as gardening, cleaning the car, or doing a walk?"). Participants answered both items by using a 4-point rating scale (1, more than once a week; 2, once a week; 3, 1 to 3 times a month; 4, hardly ever or never). Participants who did not perform any of these activities (never engaged in vigorous or moderate physical activity) were classified as "physically inactive" (1), and individuals who performed at least one vigorous or moderate physical activity were classified as "physically active" (0).

Statistical Analysis

Given that the SHARE project includes national samples from different countries and the sample design is not uniform among them, calibrated individual weights were used in the descriptive statistical analysis (for further details, see [63,64]). To compare sociodemographic, economic, and health characteristics between individuals who experience high levels of social isolation and individuals with low levels of social isolation, the chi-square test and Student *t* test were performed. We used the chi-square test to assess the interdependence between two nominal variables (eg, differences between gender according to their degree of social isolation). The sample means were also compared using Student *t* tests for independent samples (eg, differences in mean age according to their degree of social isolation). Statistical test results with $P<.05$ were considered to be significant. The results from these tests were complemented with effect size measures (Cohen d/ϕ). The interpretation of the results was based on Cohen [65].

To analyze the relationship between internet use and social isolation, logistic regressions were conducted by country using the Enter method (in which all the dependent variables are inserted in the statistical model simultaneously). The logistic regression is a mathematical model that give odds ratios (ORs) that are adjusted for other covariates (including confounders) [66]. Thus, 18 separate analyses were performed (1 for each country). Weights were not used in the logistic regressions. These analyses were performed using SPSS software (version 25; IBM Corp).

Results

Eastern and Southern Europe are the opposite of Northern and Central Europe in terms of social isolation (Table 1). In fact, the percentage of individuals aged 50 years and older who are socially isolated is higher in Eastern and Southern Europe (with the exception of Portugal) compared to the two other European regions. An analysis by country confirmed that the highest weighted percentage of people who experience a high level of social isolation is found in Poland (n=463; 36.8%), and the lowest is found in Denmark (n=321; 11.8%).

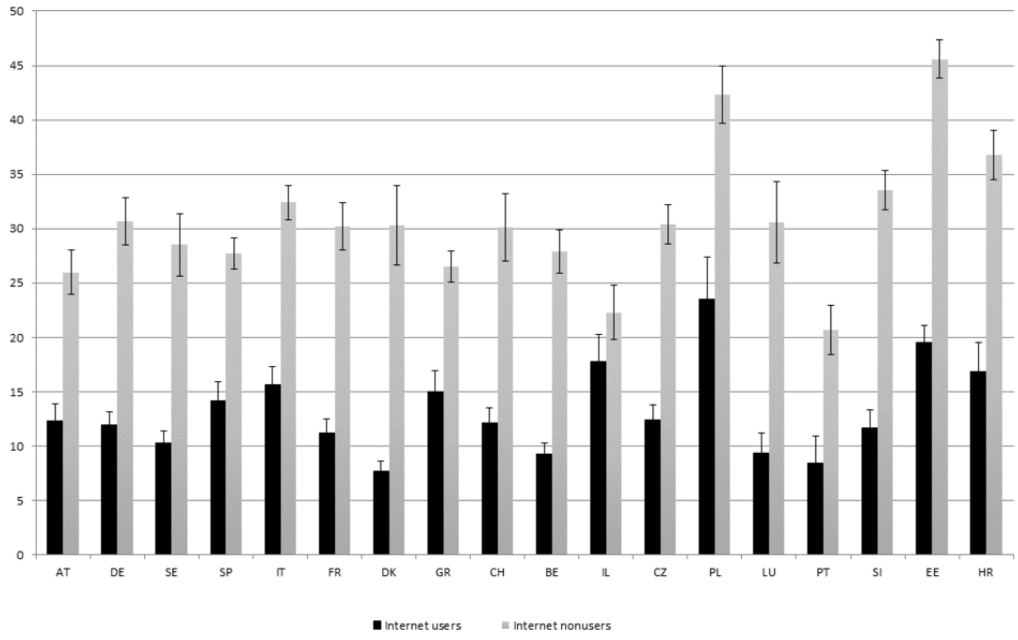
With regard to internet use, the countries of Northern and Central Europe also stand out in terms of recording the highest percentages of users, with Denmark once again showing the highest rate of use of technology (n=3053; 81.9%). In contrast, the countries of Eastern and Southern Europe have the lowest rates of internet use, with Croatia standing out as a country where fewer individuals use the internet (n=715; 27.2%).

In all the countries analyzed, high levels of social isolation are clearly more common among individuals who do not use the internet as compared to those using this technology (Figure 1). Indeed, while on average, in the group of internet users, only 12.9% experience high levels of social isolation, in the group of nonusers, 31.4% are in this situation.

Table 1. Individuals in the high isolation group (n=11,614) and internet users (n=32,399) by country (N unweighted). Source: Survey of Health, Ageing and Retirement in Europe wave 6, version 7.0.0; weighted data.

Country	High social isolation group, n (%)	Internet users, n (%)
Poland	463 (36.8)	447 (28.1)
Estonia	1408 (33.0)	2490 (47.6)
Croatia	486 (31.2)	715 (27.2)
Italy	914 (26.6)	1698 (33.6)
Slovenia	726 (24.5)	1564 (40.9)
Greece	1010 (23.3)	1265 (27.9)
Spain	946 (22.6)	1655 (37.5)
Czech Republic	1028 (21.2)	2268 (50.9)
Israel	307 (20.0)	949 (50.0)
Germany	552 (19.8)	2592 (57.6)
Austria	602 (18.8)	1606 (52.6)
France	739 (18.8)	2178 (59.8)
Luxembourg	184 (18.0)	955 (59.6)
Switzerland	395 (17.2)	1912 (71.4)
Portugal	231 (16.9)	451 (30.8)
Belgium	891 (15.8)	3684 (65.1)
Sweden	411 (14.1)	2917 (78.6)
Denmark	321 (11.8)	3053 (81.9)

Figure 1. Percentages of users and non–internet users in the high isolation group by country.



In Table 2, we can observe the sociodemographic, economic, and health characteristics of the individuals who participated in this study, according to their degree of social isolation. The group of individuals who experience high levels of social isolation includes a majority of older female adults, who have on average fewer years of schooling in relation to their counterparts experiencing low levels of social isolation. Similarly, a higher percentage of the more socially isolated individuals reported having a negative financial situation.

With regard to social networks, the more isolated individuals are more geographically distant from members of their network

than their less isolated counterparts. Likewise, more socially isolated individuals also experience greater feelings of loneliness.

Finally, with regard to mental and physical health, individuals experiencing high levels of social isolation also report more depressive symptoms and physical inactivity than their less isolated peers.

In Table 3, we can observe a negative association between internet use and social isolation, after controlling for a set of sociodemographic, economic, and health characteristics of the individuals, described in the literature as determinants of social isolation. Indeed, in most of the countries analyzed, internet use relates to a decreased likelihood of high levels of social isolation. Denmark is the country where internet users are less likely to experience high levels of social isolation compared to nonusers. In this country, internet users are 72.2% less likely (OR 0.278, 95% CI 0.196-0.396) to be in a situation of high social isolation. Similarly, in France, Sweden, Luxembourg, Germany, Czech Republic, Estonia, Switzerland, Austria, Italy, Slovenia, and Spain, internet users are less likely to experience high levels of social isolation than nonusers. In Portugal, Poland, Greece, and Croatia, internet users were not less likely to experience high social isolation. Outside the European context, more specifically in Israel, internet users were 42.7% less likely (OR 0.573, 95% CI 0.383 to 0.856) to experience high levels of social isolation.

Table 2. Descriptive statistics of the variables studied according to degree of social isolation (N=67,173). Source: Survey of Health, Ageing and Retirement in Europe wave 6, version 7.0.0; weighted data.

Variable	High social isolation (n=11,614)	Low social isolation (n=55,559)	χ^2/t (*df*)	*P* value	Effect size (Cohen *d*/ϕ)
Sociodemographic and economic characteristics					
Age (years), mean (SD)	70.05 (11.95)	65.07 (10.26)	48.781 (1)	<.001	0.498 (small)
Sex, n (%)			1457.933 (1)	<.001	0.147 (small)
Female	8359 (67.6)	29,243 (50.3)			
Male	3255 (32.4)	26,316 (49.7)			
Schooling (years), mean (SD)	9.51 (4.26)	11.17 (4.51)	–32.922 (1)	<.001	0.344 (small)
Financial situation, n (%)			976.178 (1)	<.001	0.122 (small)
Positive	5309 (48.1)	34,471 (64.5)			
Negative	5799 (51.9)	19,669 (35.5)			
Geographical distance from social network (km), mean (SD)	3.89 (1.57)	3.06 (1.61)	58.917 (1)	<.001	0.652 (medium)
Loneliness, mean (SD)	4.66 (1.78)	3.78 (1.27)	52.414 (1)	<.001	0.540 (medium)
Health					
Exhibits depressive symptoms, n (%)			1327.652 (1)	<.001	0.144 (small)
Yes	4697 (43.0)	12,751 (25.9)			
No	6713 (57.0)	39,556 (74.1)			
Limitations in activities of daily living, n (%)			662.02 (1)	<.001	0.099 (trivial)
≥1	2253 (18.9)	5977 (11.0)			
None	9358 (81.1)	49,440 (89.0)			
Physical activity, n (%)			925.115 (1)	<.001	0.117 (small)
Inactive	2434 (23.5)	5929 (12.2)			
Active	9178 (76.5)	49,473 (87.8)			

Table 3. Statistics related to the importance of the internet as a determinant of a high level of isolation in individuals aged ≥50 years. Data source: Survey of Health, Ageing and Retirement in Europe wave 6, version 7.0.0 (unweighted).

Country[a]	B[b]	aOR[c] (95% CI)	*P* value
Austria	–0.616	0.540 (0.419-0.697)	<.001
Germany	–0.797	0.451 (0.352-0.576)	<.001
Sweden	–0.842	0.431 (0.316-0.588)	<.001
Spain	–0.328	0.720 (0.558-0.931)	.01
Italy	–0.554	0.574 (0.415-0.795)	.001
France	–0.909	0.403 (0.312-0.520)	<.001
Denmark	–1.279	0.278 (0.196-0.396)	<.001
Greece	–0.062	0.940 (0.717-1.232)	.65
Switzerland	–0.675	0.509 (0.376-0.690)	<.001
Belgium	–0.808	0.448 (0.356-0.557)	<.001
Israel	–0.557	0.573 (0.383-0.856)	.007
Czech Republic	–0.795	0.452 (0.367-0.557)	<.001
Poland	–0.166	0.847 (0.521-1.377)	.50
Luxembourg	–0.802	0.448 (0.275-0.732)	.001
Portugal	0.351	1.421 (0.862-2.342)	.168
Slovenia	–0.500	0.606 (0.430-0.856)	.004
Estonia	–0.788	0.455 (0.371-0.557)	<.001
Croatia	–0.300	0.741 (0.529-1.038)	.08

[a]Unweighted n values: Austria, 2523; Germany, 3853; Sweden, 3205; Spain, 4079; Italy, 3297; France, 2904; Denmark, 3110; Greece, 4128; Switzerland, 2329; Belgium, 3999; Israel, 1242; Czech Republic, 3941; Poland, 1161; Luxembourg, 1394; Portugal, 1284; Slovenia, 2523; Estonia, 4154; Croatia, 2274.

[b]B: Standardized Coefficients

[c]aOR: adjusted odds ratio from the logistic regression with adjustment for age, gender, years of schooling, self-perception of financial stress, limitations to activities of daily life, EURO-D score, physical inactivity, geographical distance from social network, and loneliness.

Discussion

Principal Findings

The aim of this study was to analyze the relationship between internet use and social isolation of individuals aged ≥50 years. The results evidence that in most of the European countries analyzed, as well as in Israel, use of the internet by adults aged ≥50 years is related to a decreased likelihood of high levels of social isolation. Thus, this study contributes to the open debate in literature [37,50] for which other surveys present contradictory results. This could be related to the results of other studies [11,20,67] that suggest that the internet may facilitate communication and therefore enable individuals to maintain important contacts [36,39,68]. The internet may therefore contribute to less social isolation, even in less favorable contexts [12], such as when older people live at a greater geographical distance from elements of their social networks following retirement, migration, or a change in residence for other reasons [2,12-15]. However, the impact of ICT cannot be explained solely in terms of the fact that it creates opportunities for communication. The literature reveals that it counters social isolation by enabling people to obtain social support and motivating individuals to participate more in activities that interest them, because it contributes to self-confidence [37] and facilitates access to services [69].

In this study, it was also possible to verify that in the European context, countries in Eastern and Southern Europe had the highest percentages of individuals experiencing high levels of social isolation. This result is consistent with results in other studies comparing the northern European countries with southern European countries, which indicate the existence of greater social isolation in the southern countries [70]. In line with other research [3,10,11], this study concludes that high levels of social isolation are more common in older female individuals [71]. This finding contradicts the results of other research that identifies men as being more isolated than women [13,18]. The results also reveal that the most isolated individuals also have more financial difficulties, as noted in other research [3,13,18], and they experience higher levels of loneliness, as also noted in other research [19]. In this study, individuals in situations of high social isolation also reported being more geographically distant from their social network. The impact of increased geographical distance on establishing social interactions at an older age was previously underlined by other surveys showing that one of the main reasons for losing elements of adult social networks is a change in residence of these individuals [72]. In terms of health, people experiencing high levels of social

isolation also report being frailer, both physically and mentally [3,18].

However, as previously mentioned, the main contribution of the study is that after controlling for the impact of these variables, internet use is associated with lower risk of social isolation.

Limitations

This study has several limitations. The main limitation is that a single item was used to measure overall internet use. Nevertheless, a yes/no response to whether someone regularly uses the internet has frequently been used to assess internet use by older people [25,45,51,67]. In the same vein, this measure was used in the SHARE project. Nonetheless, considering that the impact of the internet in the social sphere depends on the type of activities conducted on the web [73-75], it is important for future studies to consider the impact that different uses of the internet may have on social isolation. Furthermore, future research should consider longitudinal analyses to explore causality.

Conclusion

The results of this research contribute to the scientific debate about the relationship between internet use and social isolation, showing that even after controlling for the main determinants of social isolation, the use of the internet is related to lower levels of isolation in several countries.

These results indicate the importance of developing public policies in Europe aimed at increasing rates of internet use as a way to ensure e-inclusion and prevent social isolation at an older age.

Conflicts of Interest

None declared.

References

1. de Jong Gierveld J, Havens B. Cross-national Comparisons of Social Isolation and Loneliness: Introduction and Overview. Can. J. Aging 2010 Mar 31;23(2):109-113. [doi: 10.1353/cja.2004.0021]
2. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. In: Proc Natl Acad Sci U S A. 2013 Apr 09 Presented at: PNAS; 2013; Berkeley p. 5797-5801 URL: http://www.pnas.org/lookup/pmidlookup?view=long&pmid=23530191 [doi: 10.1073/pnas.1219686110]
3. Shankar A, McMunn A, Banks J, Steptoe A. Loneliness, social isolation, and behavioral and biological health indicators in older adults. Health Psychol 2011 Jul;30(4):377-385. [doi: 10.1037/a0022826] [Medline: 21534675]
4. Shankar A, Rafnsson SB, Steptoe A. Longitudinal associations between social connections and subjective wellbeing in the English Longitudinal Study of Ageing. Psychol Health 2015;30(6):686-698. [doi: 10.1080/08870446.2014.979823] [Medline: 25350585]
5. Heffner KL, Waring ME, Roberts MB, Eaton CB, Gramling R. Social isolation, C-reactive protein, and coronary heart disease mortality among community-dwelling adults. Soc Sci Med 2011 May;72(9):1482-1488 [FREE Full text] [doi: 10.1016/j.socscimed.2011.03.016] [Medline: 21492978]
6. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci 2015 Mar;10(2):227-237. [doi: 10.1177/1745691614568352] [Medline: 25910392]
7. Leigh-Hunt N, Bagguley D, Bash K, Turner V, Turnbull S, Valtorta N, et al. An overview of systematic reviews on the public health consequences of social isolation and loneliness. Public Health 2017 Nov;152:157-171. [doi: 10.1016/j.puhe.2017.07.035] [Medline: 28915435]
8. Cantarero-Prieto D, Pascual-Sáez M, Blázquez-Fernández C. Social isolation and multiple chronic diseases after age 50: A European macro-regional analysis. PLoS One 2018;13(10):e0205062 [FREE Full text] [doi: 10.1371/journal.pone.0205062] [Medline: 30356322]
9. Schrempft S, Jackowska M, Hamer M, Steptoe A. Associations between social isolation, loneliness, and objective physical activity in older men and women. BMC Public Health 2019 Jan 16;19(1):74 [FREE Full text] [doi: 10.1186/s12889-019-6424-y] [Medline: 30651092]
10. Iliffe S, Kharicha K, Harari D, Swift C, Gillmann G, Stuck AE. Health risk appraisal in older people 6: factors associated with self-reported poor vision and uptake of eye tests in older people. BMC Fam Pract 2013 Sep 03;14:130 [FREE Full text] [doi: 10.1186/1471-2296-14-130] [Medline: 24006949]
11. Lelkes O. Happier and less isolated: internet use in old age. J Poverty Soc Justice 2013 Feb 28;21(1):33-46. [doi: 10.1332/175982713X664047]
12. Antonucci T, Ajrouch K, Manalel J. Social Relations and Technology: Continuity, Context, and Change. Innovation in Aging 2017;1(3):9. [doi: 10.1093/geroni/igx029]
13. Cudjoe T, Roth D, Szanton S, Wolff J, Boyd C, Thorpe R. The Epidemiology of Social Isolation: National Health and Aging Trends Study. J Gerontol B Psychol Sci Soc Sci 2020 Jan 01;75(1):107-113 [FREE Full text] [doi: 10.1093/geronb/gby037] [Medline: 29590462]

14. Cornwell B, Laumann EO. The health benefits of network growth: new evidence from a national survey of older adults. Soc Sci Med 2015 Jan;125:94-106 [FREE Full text] [doi: 10.1016/j.socscimed.2013.09.011] [Medline: 24128674]
15. Beckenhauer JIL, Armstrong J. Exploring Relationships Between Normative Aging, Technology, and Communication. Marriage Fam Rev 2009 Oct 21;45(6-8):825-844. [doi: 10.1080/01494920903224418]
16. Pavela G. Functional Status and Social Contact Among Older Adults. Res Aging 2015 Nov;37(8):815-836 [FREE Full text] [doi: 10.1177/0164027514566091] [Medline: 25651594]
17. DiNapoli EA, Wu B, Scogin F. Social isolation and cognitive function in Appalachian older adults. Res Aging 2014 Mar;36(2):161-179. [doi: 10.1177/0164027512470704] [Medline: 25650688]
18. Menec VH, Newall NE, Mackenzie CS, Shooshtari S, Nowicki S. Examining individual and geographic factors associated with social isolation and loneliness using Canadian Longitudinal Study on Aging (CLSA) data. PLoS One 2019;14(2):e0211143 [FREE Full text] [doi: 10.1371/journal.pone.0211143] [Medline: 30707719]
19. Shankar A, McMunn A, Demakakos P, Hamer M, Steptoe A. Social isolation and loneliness: Prospective associations with functional status in older adults. Health Psychol 2017 Feb;36(2):179-187. [doi: 10.1037/hea0000437] [Medline: 27786518]
20. Khosravi P, Rezvani A, Wiewiora A. The impact of technology on older adults' social isolation. Comput Hum Behav 2016 Oct;63:594-603. [doi: 10.1016/j.chb.2016.05.092]
21. Antonucci TC, Ajrouch KJ, Birditt KS. The convoy model: explaining social relations from a multidisciplinary perspective. Gerontologist 2014 Feb;54(1):82-92. [doi: 10.1093/geront/gnt118] [Medline: 24142914]
22. Carstensen LL. Evidence for a Life-Span Theory of Socioemotional Selectivity. Curr Dir Psychol Sci 2016 Jun 22;4(5):151-156. [doi: 10.1111/1467-8721.ep11512261]
23. Charles ST, Carstensen LL. Social and emotional aging. Annu Rev Psychol 2010;61:383-409 [FREE Full text] [doi: 10.1146/annurev.psych.093008.100448] [Medline: 19575618]
24. Khan R, Antonucci T. Convoys over the life course: Attachment, roles,social support. In: Baltes B, Brim O, editors. Life-span development and behavior. New York: Academic Press; 1980.
25. König R, Seifert A, Doh M. Internet use among older Europeans: an analysis based on SHARE data. Univ Access Inf Soc 2018 Jan 19;17(3):621-633. [doi: 10.1007/s10209-018-0609-5]
26. Silva P, Matos AD, Martinez-Pecino R. E-inclusion: Beyond individual socio-demographic characteristics. PLoS One 2017;12(9):e0184545 [FREE Full text] [doi: 10.1371/journal.pone.0184545] [Medline: 28910358]
27. Selwyn N, Gorard S, Furlong J, Madden L. Older adults' use of information and communications technology in everyday life. Ageing Soc 2003 Sep 02;23(5):561-582. [doi: 10.1017/S0144686X03001302]
28. Matthews K, Nazroo J, Marshall A. Digital inclusion in later life: cohort changes in internet use over a ten-year period in England. Ageing Soc 2018 Aug 02;39(9):1914-1932. [doi: 10.1017/s0144686x18000326]
29. Schlomann A, Seifert A, Zank S, Rietz C. Assistive Technology and Mobile ICT Usage Among Oldest-Old Cohorts: Comparison of the Oldest-Old in Private Homes and in Long-Term Care Facilities. Res Aging 2020;42(5-6):163-173. [doi: 10.1177/0164027520911286] [Medline: 32167019]
30. Anderberg P, Eivazzadeh S, Berglund JS. A Novel Instrument for Measuring Older People's Attitudes Toward Technology (TechPH): Development and Validation. J Med Internet Res 2019 May 23;21(5):e13951 [FREE Full text] [doi: 10.2196/13951] [Medline: 31124467]
31. Knapova L, Klocek A, Elavsky S. The Role of Psychological Factors in Older Adults' Readiness to Use eHealth Technology: Cross-Sectional Questionnaire Study. J Med Internet Res 2020 May 28;22(5):e14670 [FREE Full text] [doi: 10.2196/14670] [Medline: 32348251]
32. Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. Cyberpsychol Behav Soc Netw 2015 May;18(5):268-272. [doi: 10.1089/cyber.2014.0549] [Medline: 25919967]
33. Seçkin G. Using Hierarchical Multivariate Analysis to Examine Interactions Between Alternative Appraisals of Cancer and Virtual Health Support Among Middle-Aged and Older Internet Users. Res Aging 2012 Oct 29;35(6):688-709. [doi: 10.1177/0164027512460692]
34. Jones RB, Ashurst EJ, Atkey J, Duffy B. Older people going online: its value and before-after evaluation of volunteer support. J Med Internet Res 2015 May 18;17(5):e122 [FREE Full text] [doi: 10.2196/jmir.3943] [Medline: 25986724]
35. Silva P, Matos AD, Martinez-Pecino R. Can the internet reduce the loneliness of 50+ living alone? Inf Commun Soc 2020 May 12:1-17. [doi: 10.1080/1369118x.2020.1760917]
36. Silva P, Delerue Matos A, Martinez-Pecino R. Confidant Network and Quality of Life of Individuals Aged 50+: The Positive Role of Internet Use. Cyberpsychol Behav Soc Netw 2018 Nov;21(11):694-702. [doi: 10.1089/cyber.2018.0170] [Medline: 30335512]
37. Chen YR, Schulz PJ. The Effect of Information Communication Technology Interventions on Reducing Social Isolation in the Elderly: A Systematic Review. J Med Internet Res 2016 Jan 28;18(1):e18 [FREE Full text] [doi: 10.2196/jmir.4596] [Medline: 26822073]
38. Cotton S, Anderson W, McCullough B. The impact of ICT use on loneliness and contact with others among older adults. Gerontechnology 2012;11(2):1-8. [doi: 10.4017/gt.2012.11.02.378.00]
39. Russell C, Campbell A, Hughes I. Ageing, social capital and the Internet: findings from an exploratory study of Australian 'silver surfers'. Australas J Ageing 2008 Jun;27(2):78-82. [doi: 10.1111/j.1741-6612.2008.00284.x] [Medline: 18713197]

40. Chopik WJ. The Benefits of Social Technology Use Among Older Adults Are Mediated by Reduced Loneliness. Cyberpsychol Behav Soc Netw 2016 Sep;19(9):551-556 [FREE Full text] [doi: 10.1089/cyber.2016.0151] [Medline: 27541746]
41. Zaine I, Frohlich DM, Rodrigues KRDH, Cunha BCR, Orlando AF, Scalco LF, et al. Promoting Social Connection and Deepening Relations Among Older Adults: Design and Qualitative Evaluation of Media Parcels. J Med Internet Res 2019 Oct 03;21(10):e14112 [FREE Full text] [doi: 10.2196/14112] [Medline: 31584001]
42. Sum S, Mathews RM, Pourghasem M, Hughes I. Internet use as a predictor of sense of community in older people. Cyberpsychol Behav 2009 Apr;12(2):235-239. [doi: 10.1089/cpb.2008.0150] [Medline: 19250013]
43. Cotten SR, Anderson WA, McCullough BM. Impact of internet use on loneliness and contact with others among older adults: cross-sectional analysis. J Med Internet Res 2013 Feb 28;15(2):e39 [FREE Full text] [doi: 10.2196/jmir.2306] [Medline: 23448864]
44. Willard S, van Rossum E, Spreeuwenberg M, de Witte L. Perceived Impact of an Online Community Care Platform for Dutch Older Adults on Local Participation, Informal Caregiving, and Feelings of Connectedness: Pretest-Posttest Observational Study. J Med Internet Res 2020 Dec 01;22(12):e20304 [FREE Full text] [doi: 10.2196/20304] [Medline: 33258779]
45. Hogeboom DL, McDermott RJ, Perrin KM, Osman H, Bell-Ellison BA. Internet Use and Social Networking Among Middle Aged and Older Adults. Educ Gerontol 2010 Jan 05;36(2):93-111. [doi: 10.1080/03601270903058507]
46. Loges W, Jung J. Exploring the Digital Divide: Internet Connectedness and Age. Communication Research 2001;28(4):536-562. [doi: 10.1177/009365001028004007]
47. Meshi D, Cotten S, Bender A. Problematic Social Media Use and Perceived Social Isolation in Older Adults: A Cross-Sectional Study. Gerontology 2020;66(2):160-168. [doi: 10.1159/000502577] [Medline: 31522178]
48. Nie N, Hillygus S, Erbring L. Internet Use, Interpersonal Relations, and Sociability: A Time Diary Study. In: Wellman B, Haythornthwaite C, editors. The Internet in Everyday Life. Oxford: Blackwell Publishers Ltd; 2002:215-243.
49. Beneito-Montagut R, Cassián-Yde N, Begueria A. What do we know about the relationship between internet-mediated interaction and social isolation and loneliness in later life? QAOA 2018 Mar 12;19(1):14-30. [doi: 10.1108/QAOA-03-2017-0008]
50. Baker S, Warburton J, Waycott J, Batchelor F, Hoang T, Dow B, et al. Combatting social isolation and increasing social participation of older adults through the use of technology: A systematic review of existing evidence. Australas J Ageing 2018 Sep;37(3):184-193. [doi: 10.1111/ajag.12572] [Medline: 30022583]
51. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among retired older adults in the United States: a longitudinal analysis. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):763-771. [doi: 10.1093/geronb/gbu018] [Medline: 24671896]
52. Malter F, Börsch-Supan A. SHARE Wave 6: Panel Innovations and collecting Dried Blood Spots. Munich: Munich Center for the Economics of Aging (MEA); 2017.
53. Shankar A, Hamer M, McMunn A, Steptoe A. Social isolation and loneliness: relationships with cognitive function during 4 years of follow-up in the English Longitudinal Study of Ageing. Psychosom Med 2013 Feb;75(2):161-170. [doi: 10.1097/PSY.0b013e31827f09cd] [Medline: 23362501]
54. Kobayashi LC, Steptoe A. Social Isolation, Loneliness, and Health Behaviors at Older Ages: Longitudinal Cohort Study. Ann Behav Med 2018 May 31;52(7):582-593 [FREE Full text] [doi: 10.1093/abm/kax033] [Medline: 29860361]
55. Stockwell S, Stubbs B, Jackson SE, Fisher A, Yang L, Smith L. Internet use, social isolation and loneliness in older adults. Ageing Soc 2020 May 19;41(12):2723-2746. [doi: 10.1017/s0144686x20000550]
56. Jackson SE, Firth JA, Firth J, Veronese N, Gorely T, Grabovac I, et al. Social isolation and physical activity mediate associations between free bus travel and wellbeing among older adults in England. J Transp Health 2019 Jun;13:274-284. [doi: 10.1016/j.jth.2019.03.006]
57. Malter F, Börsch-Supan A. Share Wave 4 Innovations & Methodology. Munich: Munich Center for the Economics of Aging (MEA); 2013.
58. Prince MJ, Reischies F, Beekman AT, Fuhrer R, Jonker C, Kivela SL, et al. Development of the EURO-D scale-a European, Union initiative to compare symptoms of depression in 14 European centres. Br J Psychiatry 1999 Apr;174:330-338. [Medline: 10533552]
59. Schwartz E, Litwin H. Are newly added and lost confidants in later life related to subsequent mental health? Int Psychogeriatr 2017 Dec;29(12):2047-2057 [FREE Full text] [doi: 10.1017/S1041610217001338] [Medline: 28714428]
60. Guerra M, Ferri C, Llibre J, Prina AM, Prince M. Psychometric properties of EURO-D, a geriatric depression scale: a cross-cultural validation study. BMC Psychiatry 2015 Feb 05;15:12 [FREE Full text] [doi: 10.1186/s12888-015-0390-4] [Medline: 25652111]
61. Litwin H, Stoeckel K, Schwartz E. Social networks and mental health among older Europeans: are there age effects? Eur J Ageing 2015;12(4):299-309 [FREE Full text] [doi: 10.1007/s10433-015-0347-y] [Medline: 28804362]
62. Mehrbrodt T, Gruber S, Wagner M. Scales and Multi-Item Indicators Internet. 2017. URL: http://www.share-project.org/fileadmin/pdf_documentation/SHARE_Scales_and_Multi-Item_Indicators.pdf [accessed 2020-01-30]

63. Klevmarken A, Hesselius P, Swensson B. The SHARE Sampling ProceduresCalibrated Design Weights. In: Börsch-Supan A, Jürges H, editors. The Survey of Health, Aging, and Retirement in Europe – Methodology. Mannheim: Research Institute for the Economics of Aging (MEA); 2005:28-69.
64. Lynn P, De Luca G, Ganninger M. Sample Design in SHARE Wave Four. In: Malter F, Börsch-supan A, editors. SHARE Wave 4: innovations & Methodology. SHARE Wave 4 Innov Methodol München: MEA. Max Planck Institute for Social Law and Social Policy; 2013:74-123.
65. Cohen J. The t test for means. In: Statistical Power Analysis for the Behavioral Sciences. New York: Routledge; 1988:19-74.
66. Pourhoseingholi MA, Baghestani AR, Vahedi M. How to control confounding effects by statistical analysis. Gastroenterol Hepatol Bed Bench 2012;5(2):79-83 [FREE Full text] [Medline: 24834204]
67. Cotten SR, Ford G, Ford S, Hale TM. Internet use and depression among retired older adults in the United States: a longitudinal analysis. J Gerontol B Psychol Sci Soc Sci 2014 Sep;69(5):763-771. [doi: 10.1093/geronb/gbu018] [Medline: 24671896]
68. Martinez-Pecino R, Delerue Matos A, Silva P. Portuguese older people and the Internet: Interaction, uses, motivations, and obstacles. Communications 2013;38:331-346 [FREE Full text] [doi: 10.1515/commun-2013-0020]
69. Berg T, Winterton R, Petersen M, Warburton J. 'Although we're isolated, we're not really isolated': The value of information and communication technology for older people in rural Australia. Australas J Ageing 2017 Dec;36(4):313-317. [doi: 10.1111/ajag.12449] [Medline: 28639293]
70. Vozikaki M, Papadaki A, Linardakis M, Philalithis A. Social isolation and well-being among older adults in Europe. Archives of Hellenic Medicine 2018;35:506-519 [FREE Full text]
71. Locher JL, Ritchie CS, Roth DL, Baker PS, Bodner EV, Allman RM. Social isolation, support, and capital and nutritional risk in an older sample: ethnic and gender differences. Soc Sci Med 2005 Feb;60(4):747-761 [FREE Full text] [doi: 10.1016/j.socscimed.2004.06.023] [Medline: 15571893]
72. Cornwell B, Schumm LP, Laumann EO, Kim J, Kim Y. Assessment of social network change in a national longitudinal survey. J Gerontol B Psychol Sci Soc Sci 2014 Nov;69 Suppl 2:S75-S82 [FREE Full text] [doi: 10.1093/geronb/gbu037] [Medline: 25360026]
73. Hampton K, Her E, Rainie L. Social Isolation and New Technology How the internet and mobile phones impact Americans' social networks.: Pew Internet & American Life Project; 2009. URL: http://www.site.uottawa.ca/~stan/csi2911/Pew.pdf [accessed 2020-01-01]
74. Zhao S. Do Internet Users Have More Social Ties? A Call for Differentiated Analyses of Internet Use. J Comp Mediated Comm 2006 Apr;11(3):844-862. [doi: 10.1111/j.1083-6101.2006.00038.x]
75. Chen W. Internet Use, Online Communication, and Ties in Americans' Networks. Soc Sci Comput Rev 2013 Mar 15;31(4):404-423. [doi: 10.1177/0894439313480345]

Abbreviations

OR: odds ratio
R-UCLA: Revised UCLA Loneliness Scale
SHARE: Survey of Health, Ageing and Retirement in Europe

Edited by R Kukafka; submitted 20.05.20; peer-reviewed by A Stuart, T Wieringa; comments to author 27.10.20; revised version received 03.12.20; accepted 23.09.21; published 03.01.22.

Please cite as:
Silva P, Delerue Matos A, Martinez-Pecino R
The Contribution of the Internet to Reducing Social Isolation in Individuals Aged 50 Years and Older: Quantitative Study of Data From the Survey of Health, Ageing and Retirement in Europe
J Med Internet Res 2022;24(1):e20466
URL: https://www.jmir.org/2022/1/e20466
doi:10.2196/20466
PMID:34982040

Original Paper

The Digital Divide and Seeking Health Information on Smartphones in Asia: Survey Study of Ten Countries

Xiaohui Wang[1], PhD; Jingyuan Shi[2], PhD; Kwan Min Lee[3], PhD

[1]Department of Media and Communication, City University of Hong Kong, Hong Kong, Hong Kong

[2]Department of Communication Studies, Hong Kong Baptist University, Hong Kong, Hong Kong

[3]Wee Kim Wee School of Communication and Information, Nanyang Technological University, Singapore, Singapore

Corresponding Author:
Kwan Min Lee, PhD
Wee Kim Wee School of Communication and Information
Nanyang Technological University
Room 03-44, WKWSCI Building
31 Nanyang Link
Singapore, 637718
Singapore
Phone: 65 63168816
Email: kwanminlee@ntu.edu.sg

Abstract

Background: Although recent developments in mobile health have elevated the importance of how smartphones empower individuals to seek health information, research investigating this phenomenon in Asian countries has been rare.

Objective: The goal of our study was to provide a comprehensive profile of mobile health information seekers and to examine the individual- and country-level digital divide in Asia.

Methods: With survey data from 10 Asian countries (N=9086), we ran multilevel regression models to assess the effect of sociodemographic factors, technological factors, and country-level disparities on using smartphones to seek health information.

Results: Respondents who were women (β=.13, *P*<.001), parents (β=.16, *P*<.001), employed (β=.08, *P*=.002), of higher social status (β=.08, *P*<.001), and/or from countries with low health expenditures (β=.19, *P*=.02) were more likely to use smartphones to seek health information. In terms of technological factors, technology innovativeness (β=.10, *P*<.001) and frequency of smartphone use (β=.42, *P*<.001) were important factors of health information seeking, whereas the effect of online information quality was marginal (β=–.04, *P*<.001).

Conclusions: Among smartphone users in Asia, health information seeking varies according to individuals' socioeconomic status, their innovativeness toward technology, and their frequency of smartphone use. Although smartphones widen the digital divide among individuals with different socioeconomic status, they also bridge the divide between countries with varying health expenditures. Smartphones appear to be a particularly useful complement to manage health in developing countries.

(J Med Internet Res 2022;24(1):e24086) doi:10.2196/24086

KEYWORDS

smartphone; health information seeking; Asia; user profile; digital divide

Introduction

Background

With the development of mobile technology and the spread of its use, smartphones have facilitated several positive health-related outcomes, including the increased provision of psychological interventions [1], improved access to health services, and reductions in various forms of social inequality [2] in populations worldwide. In that context, mobile health (mHealth) refers to public health or medical practices that involve using mobile information communication technologies (ICTs) to seek out health information, communicate with health care professionals, and monitor personal health [3]. Since mHealth provides individuals with ubiquitous access to timely health services at low cost, it can play a major role in bridging inequalities in digital access to health care services [4,5].

In particular, the act of seeking health-related information online has received substantial research attention, but mainly in western

countries. For example, analyses of the data from several surveys on the topic administered in the United States and Germany [6,7] have provided valuable insights into the determinants and outcomes of such information-seeking behavior in those populations. By contrast, research investigating such behavior in Asian countries has been sorely limited, especially with respect to health information seeking on smartphones [8].

To address this gap in the literature, we designed this study to investigate the factors contributing to mHealth information seeking (MHIS) to profile such information seekers in Asia using data from 10 Asian countries. In effect, our comprehensive profiling stands to help service providers understand the mHealth market from the perspective of users. We also sought to examine the extent to which individual- and country-level digital divides exist in MHIS across various Asian countries. To the best of our knowledge, this project represents by far the largest multinational survey on smartphone users in Asia. Thus, its findings are expected to provide theoretical insights into both MHIS and its practical implications, all of which may support efforts to bridge the digital divide in Asia.

Profiling mHealth Information Seekers

The primary objective of this study was to investigate factors related to MHIS to ultimately profile such information seekers in Asia. Previous research has examined correlates of using mHealth that can be classified into four groups: consumers' motivations and needs for health information [9,10]; driving factors behind adopting mHealth according to theories about technology and health behavior (eg, technology acceptance model, diffusion of innovation theory, and comprehensive model of information seeking) [8,11,12]; sociodemographic features of mHealth users, including their age, gender, education, and income [13-16]; and boundary conditions (eg, cultural context) of theoretical models applied in examining the use of smartphones for health-related purposes [17-19].

Zhao et al [11] recently performed a meta-analysis of 35 empirical studies on mHealth, the overall results of which diverged depending on the context studied and the characteristics of the sample. Their meta-analysis also revealed that the findings of most studies have been based on data from small samples in single-culture contexts, whereas few studies have used multinational data from diverse cultural contexts. Moreover, Wang et al [8] found that technological factors were important yet neglected factors in determining individuals' online health information–seeking behaviors. Thus, in response to these gaps, we profiled mHealth information seekers in Asia according to their sociodemographic characteristics as well as perceptions of technology.

Parallel to the rise of mHealth technologies, discussions about the quality of health information available online have also intensified. From the perspective of information systems, a high level of *information quality*—defined as the credibility and reliability of information in terms of relevance, sufficiency, accuracy, and timeliness [20]—promotes the use of information systems [21]. At the same time, studies have highlighted users' concerns with the quality of health information available online, which could hinder online information seeking about health-related topics [22]. Nevertheless, few studies have examined how information quality affects the behavior of MHIS [8].

Against this trend, research has suggested that the adoption of new technology may be primarily driven by how individuals perceive its innovativeness [23]. According to the diffusion of innovation theory, *technology innovativeness*, defined as the degree to which an individual perceives a technological device to be innovative and is thus willing to experiment with using it, is an important factor for adopting technology [24]. Recently, technology innovativeness was also identified as a significant determinant of the intention to use mHealth [25]. In light of these findings, we aimed to scrutinize the effect of online information quality and technology innovativeness on MHIS.

The Digital Divide and Seeking Health Information on Smartphones

The ability to seek health-related information online has been found to enable individuals to make informed decisions about their health, provide individuals in need with disease-related social support, and help patients adhere to their medication and treatment regimens [8,19]. Nevertheless, extensive research has also revealed that not all individuals benefit equally from assessing and consuming health information online [6,7,26,27]. A critical reason for this inequality is the so-called "digital divide," a phenomenon related not only to internet access but also to the existence of a gap between people who can and cannot effectively use new communication tools (eg, smartphones) or comprehend new information [28]. Past findings have additionally suggested that smartphones can act either as a bridge or as a barrier for people in assessing health-related information, depending on their socioeconomic status (SES) [6,29-31].

Thus, in a final contribution to the literature, we also tested the digital divide in MHIS at the individual as well as national levels in Asia. To date, scholars have suggested that the global digital divide has narrowed the most for mobile phone use, likely because many developing countries have simply stopped using fixed-line communication as their access to technology has advanced [32]. Therefore, we expected that with smartphones, residents in less developed countries can easily access the internet and health information, which may in turn reduce the inequality in MHIS in those countries.

Methods

Procedure and Participants

For our sample, 9086 adults in 10 Asian countries (China, India, Indonesia, Thailand, the Philippines, Malaysia, South Korea, Japan, Vietnam, and Singapore) were recruited in 2016. In line with the procedures of cross-cultural research [33,34], the data were collected from major cities in all countries using synchronous data collection between June 2016 and October 2016. The questionnaire developed for this purpose was translated into each country's dominant language using standard translation and back-translation, after which it was distributed in each country through an online survey company. To participate, individuals had to have a smartphone and be between 18 and 55 years of age. The total sample comprised at least 800

respondents from each country who were evenly distributed by age and gender in order to fully capture the situation of each stratum in the population. Given these methods, the Institutional Review Board at Nanyang Technological University approved the study.

The participants were adult smartphone users from 10 Asian countries: mainland China (n=1238), India (n=1238), Indonesia (n=824), Thailand (n=821), the Philippines (n=843), Malaysia (n=837), South Korea (n=858), Japan (n=804), Vietnam (n=809), and Singapore (n=814). In total, 9086 smartphone users were recruited. Participants ranged in age from 18 to 55 years, and the majority were women. Nearly three-quarters of the participants had completed college, university, and/or graduate school, and more than half were married and had at least one child. Approximately 88% of participants were employed or self-employed. Detailed descriptive statistics of the sample are given in Table 1.

Table 1. Demographic characteristics of the sample.

Characteristic	All participants (N=9086)	Frequent seekers[a] (n=6508)
Gender, n (%)		
Man	4716 (51.90)	3292 (50.58)
Woman	4370 (48.10)	3216 (49.42)
Age (years), mean (SD)	34.3 (9.11)	33.7 (8.80)
Marital status, n (%)		
Married	5369 (59.09)	3889 (59.76)
Single, divorced, separated, or widowed	3717 (40.91)	2619 (40.24)
Have child(ren), n (%)		
No children	3924 (43.19)	2526 (38.81)
At least one child	5162 (56.81)	3982 (61.19)
Level of education, n (%)		
High school or less	2580 (28.40)	1686 (25.91)
College or university	5101 (56.14)	3755 (57.70)
Graduate school or more	1405 (15.46)	1067 (16.40)
Employment status, n (%)		
Employed	7971 (87.73)	5808 (89.24)
Unemployed, homemaker, or retired	1115 (12.27)	700 (10.76)
Smartphone use, mean (SD)	4.87 (1.17)	5.16 (1.01)
Concern with online information quality, mean (SD)	5.15 (1.18)	5.23 (1.15)
Technology innovativeness, mean (SD)	4.42 (1.46)	4.68 (1.34)

[a]"Frequent seekers" reported seeking health information on their smartphones at least a few times per month.

Measures

Quantifying MHIS

Participants reported how frequently they used their smartphones to seek information about health and medical issues on a 5-point scale (1=*never*, 2=*rarely* [ie, once per month], 3=*sometimes* [ie, a few times per month], 4=*often* [ie, a few times per week], 5=*always* [ie, daily]). This item was adopted to measure participants' MHIS (mean 3.17, SD 1.23).

Demographics

Participants reported their gender (0=*man*, 1=*woman*), age, marital status (1=*married*, 0=*single, divorced, separated, or widowed*), and parental status (1=*have child(ren)*, 0=*no children*).

Objective SES

Level of education, employment status, and monthly household income were employed as indicators of objective SES [35]. Participants were asked to report their level of education (0=*high school or less*, 1=*college or university*, 2=*graduate school or more*), employment status (0=*unemployed*, 1=*employed*), and monthly income. Given that income levels vary across countries, we standardized the reported monthly income within each country to enable comparisons and analyses across the 10 countries.

Subjective SES

A scale for measuring subjective SES was employed [36], in which participants were asked to rate their perceptions of their SES on a 10-rung hierarchical scale. The bottom of the scale, where the score was 1, represented participants who perceived themselves as having the least wealth, the least education, and the least-respected jobs, or no job whatsoever, compared with

those of others. The top of the scale, where the score was 10, represented participants who perceived themselves as having the most wealth, the most education, and the most respected jobs relative to others. Participants were asked to place an "X" on the rung of the scale that they believed best reflected their SES (mean 5.84, SD 1.78).

Technological Factors

Technological factors included participants' self-reported perceptions of technology and frequency of smartphone use. Items were adapted from previous studies that involved the measurement of similar concepts [37,38].

Perceived online information quality was assessed on a 7-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Participants rated to what extent they agree with the following two items: "In general, there is less control over the quality of the content posted online" and "There is a lot of fake news online these days." The answers were recoded such that a larger number indicated a higher perceived quality of information online (mean 5.16, SD 1.22; r=0.59).

Technology innovativeness was assessed on a 7-point Likert scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Participants indicated their technology innovativeness on the following three items: "Your friends describe you as 'into the latest technology'," "You often purchase new technology before your friends," and "You consider yourself technologically sophisticated" (mean 4.43, SD 1.40; Cronbach α=.91).

The frequency of smartphone use was determined according to participants reporting how often they used their smartphones for the following activities in the past year: (1) keeping up to date with news and information, (2) using social networking sites or apps, (3) using electronic banking, and (4) playing games or watching entertaining videos. The answers were recorded on a 5-point Likert scale where 1=*never*, 2=*rarely* (once a month), 3=*sometimes* (few times a month), 4=*often* (few times a week), and 5=*always* (daily). The four items were averaged on a scale of the frequency of smartphone use (mean 3.91, SD 0.78).

Country-Level Digital Divide

The country-level digital divide was measured using the ICT Development Index (IDI) from the International Telecommunication Union database of development indicators [39], which indicates the level of the development of ICT infrastructure in a given country. The IDI generates a score from 1 to 10, and of the countries sampled, India scored the lowest (ie, 3.03) and South Korea scored the highest (ie, 8.85). Economic inequality at the country level was measured using the Gini index from the database of World Bank Development Indicators [40]. The Gini index generally ranges from 0% (*perfect equality*) to 100% (*perfect inequality*); of the countries sampled, the Philippines scored the highest (ie, 44.4%), whereas South Korea scored the lowest (ie, 31.6%). Health inequality at the country level was measured by each country's current health expenditure (CHE) per capita based on the global health expenditure database maintained by the World Health Organization [41], which is reported in purchasing power parity, ranging from 253 in India to 4563 in Japan.

Data Analysis

Multilevel linear regression models were constructed to determine the effect of individual-level characteristics and country-level inequalities in MHIS, after which analyses were performed using the lme4 package in R, an open-source program for statistical analysis. The model-building process involved three steps: (1) a univariate analysis of each variable using appropriate statistical tests (eg,t test, analysis of variance, or Pearson correlation), (2) a model including any variables whose univariate test had a P value less than .25, and (3) a two-level linear model with variances specified at the individual level and country level. In this paper, the results of multilevel linear regression analysis are presented in terms of their standardized β coefficients and model statistics. A two-sided P value of less than .05 for all tests was considered to be statistically significant.

Results

Descriptive Analysis

The sociodemographic characteristics of the participants are shown in Table 1. Of the 9086 smartphone users sampled, only 996 (10.96%) had never sought health-related information on their smartphones. By contrast, 71.63% (n=6508) reported seeking such information on their smartphones at least a few times per month. Other than seeking health-related information, most participants reported regularly using smartphones to keep themselves informed of the latest news and information (8650/9086, 95.20%), build and maintain social networks (8432/9086, 92.80%), use electronic banking (7578/9086, 83.40%), play video games (7196/9086, 79.20%), and watch entertaining videos (7196/9086, 79.20%). Generally, the participants considered themselves to be interested in technological innovation and new technology (mean 4.42, SD 1.46). Regarding their concerns over the quality of online information, more than half agreed (Likert scale>4) that there is less control over the quality of online content than content from other media outlets (6108/9086, 67.22%) and that there is a lot of fake news online (6766/9086, 74.47%).

Individual-Level Analyses

In the univariate analyses, the frequency of using smartphones to find health information was significantly higher among participants who were women (t_{9085}=4.00, P<.001), married (t_{9085}=–3.25, P=.01), employed (t_{9085}=–9.06, P<.001), parents (t_{9085}=–16.83, P<.001), and/or had a high monthly household income level (r=0.06, P<.001) and a high level of education ($F_{2,9084}$=58.0, P<.001). Moreover, the frequency of using smartphones to seek health information was positively associated with the level of subjective SES (r=0.29, P<.001) but was negatively associated with age (r=–0.09, P<.001).

In terms of technology-related factors, the frequency of using smartphones to search for health information was significantly and positively associated with the frequency of smartphone use in general (r=0.51, P<.001). Additionally, concern about online information quality (r=0.14, P<.001) was significantly and positively related to seeking health information on smartphones.

The results of the multilevel regression analysis substantially confirmed the findings of the univariate analyses (Table 2). For individual-level characteristics, individuals who frequently sought health-related information using their smartphones were more likely to be women, to be employed, to have at least one child, and to perceive themselves as having a high SES. At the same time, age, household income, and marital status were not significantly related to MHIS in the regression model. Frequencies of smartphone use and technology innovativeness were positively related to seeking health information. However, the relationship between concern over online information quality and MHIS became negative with the presence of other factors. Model statistics are summarized in Table 3.

Table 2. Multilevel regression analyses of mobile health information seeking.

Variable	Model 1	Model 2		Model 3	
		β (SE)	*P* value	β (SE)	*P* value
Intercept	–.02 (.15)	.06 (.07)	.09	–.08 (.05)	.07
Individual level					
Gender					
Woman	—[a]	Reference	Reference	Reference	Reference
Man	—	–.13 (.02)	<.001	–.13 (.02)	<.001
Age	—	–.01 (.01)	.13	.01 (.01)	.11
Marital status					
Single, divorced, separated, or widowed	—	Reference	Reference	Reference	Reference
Married	—	-.004 (.02)	.42	–.004 (.02)	.41
Education					
High school or less	—	Reference	Reference	Reference	Reference
College or university	—	.02 (.03)	.14	.02 (.02)	.13
Graduate school or more	—	.08 (.02)	.004	.07 (.03)	.009
Monthly income	—	-.02 (.01)	.06	-.02 (.01)	.06
Subjective SES[b]	—	.08 (.01)	<.001	.08 (.01)	<.001
Employment					
Unemployed	—	Reference	Reference	Reference	Reference
Employed	—	.08 (.02)	<.001	.08 (.02)	.002
Parental status					
No children	—	Reference	Reference	Reference	Reference
Have at least one child	—	.16 (.02)	<.001	.16 (.02)	<.001
Concern with online information quality	—	–.04 (.01)	<.001	–.04 (.01)	<.001
Technology innovativeness	—	.11 (.01)	<.001	.10 (.01)	<.001
Frequency of smartphone use	—	.42 (.01)	<.001	.42 (.01)	<.001
Country level					
IDI[c]	—	—	—	.06 (.09)	.26
CHE[d]	—	—	—	–.19 (.09)	.02
GINI[e]	—	—	—	–.03 (.05)	.22

[a]Not included in model.

[b]SES: socioeconomic status.

[c]IDI: Information Communications Technology Development Index.

[d]CHE: current health expenditure per capita (purchasing power parity, 2017).

[e]GINI: Gini index (World Bank estimate).

Table 3. Statistics for the multivariate regression models.

Model statistic	Model 1	Model 2	Model 3
Level 1 variance (SD)	0.88 (0.94)	0.66 (0.18)	0.66 (0.82)
Level 2 variance (SD)	0.15 (0.38)	0.03 (0.18)	0.02 (0.13)
Intraclass coefficient	N/A[a]	N/A	0.14
Log-likelihood	24685.7	22167.3	22160.1

[a]N/A: not applicable.

Country-Level Analyses

Table 4 presents the overall descriptive analysis at the country level. We found that the frequency of seeking health information using smartphones was the highest in Vietnam (mean 4.57, SD 1.83), followed by Indonesia (mean 4.31, SD 1.63), India (mean 4.06, SD 1.73), China (mean 4.00, SD 1.60), the Philippines (mean 3.96, SD 1.62), Thailand (mean 3.94, SD 1.72), Malaysia (mean 3.53, SD 1.68), South Korea (mean 3.48, SD 1.72), Singapore (mean 3.10, SD 1.54), and Japan (mean 2.29, SD 1.46). The distribution of overall smartphone use across countries showed a similar pattern to that of using such technology to seek health information. People in Vietnam (mean 5.05, SD 1.15), China (mean 5.03, SD 0.94), India (mean 4.90, SD 1.06), Thailand (mean 4.85, SD 1.05), Indonesia (mean 4.80, SD 1.00), and the Philippines (mean 4.76, SD 0.93) reported a higher frequency of using their smartphones for news and information, electronic banking, social networking, and entertainment than people in Malaysia (mean 4.44, SD 1.06), South Korea (mean 4.40, SD 1.08), Singapore (mean 4.18, SD 1.06), and Japan (mean 3.61, SD 1.20).

Figure 1 shows a scatterplot of the relationship between CHE, IDI, the Gini index, and MHIS among the 10 Asian countries. All three factors of inequality at the country level were highly related to information-seeking behavior. Countries with a low Gini index, low IDI, and low CHE were more likely to rely on smartphones as a source for their health information. However, the multilevel regression analysis (Table 2) showed that CHE was the only significant factor related to individuals' MHIS. Economic inequality and IDI were not significantly associated with MHIS in the regression model.

Table 4. Country-level statistics.

Country	Sample, n	MHIS[a] (SD)	General smartphone use (SD)	COIQ[b] (SD)	IDI[c]	CHE[d]	GINI[e]
China	1238	3.34 (1.11)	4.11 (0.67)	4.85 (1.12)	5.60	841.1	38.6
India	1238	3.39 (1.19)	4.09 (0.71)	5.36 (1.20)	3.03	253.3	35.7
Indonesia	824	3.58 (1.11)	4.03 (0.70)	5.31 (1.11)	4.33	367.9	38.1
Japan	804	2.09 (1.10)	3.11 (0.93)	4.28 (1.43)	8.43	4563	32.1
South Korea	858	2.97 (1.21)	3.68 (0.79)	4.94 (1.19)	8.85	2980	31.6
Malaysia	837	3.03 (1.18)	3.82 (0.74)	5.37 (1.06)	6.38	1139	41.0
Philippines	843	3.37 (1.12)	4.10 (0.64)	5.55 (1.10)	4.67	371.7	44.4
Singapore	814	2.75 (1.11)	3.64 (0.78)	5.23 (1.00)	8.05	4270	35.6
Thailand	821	3.29 (1.20)	4.10 (0.71)	5.35 (1.12)	5.67	670.9	36.5
Vietnam	809	3.72 (1.20)	4.18 (0.71)	5.46 (1.33)	4.43	375.6	35.5
All	9086	3.17 (1.23)	3.91 (0.80)	5.16 (1.22)	5.94	1583	36.9

[a]MHIS: mobile health information seeking.

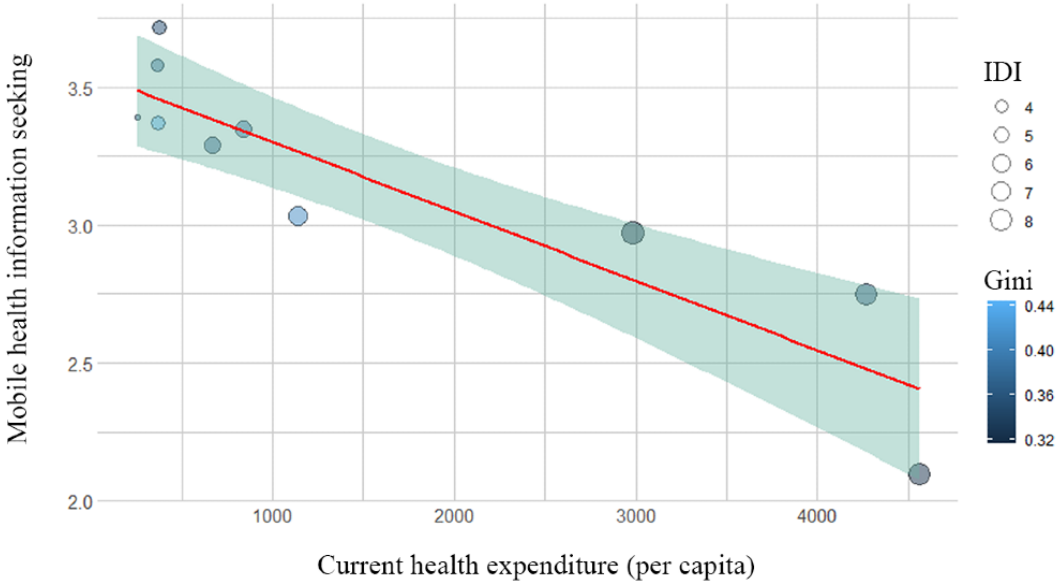
[b]COIQ: concern with online information quality.

[c]IDI: Information Communications Technology Development Index.

[d]CHE: current health expenditure per capita (purchasing power parity, 2017).

[e]GINI: Gini index (World Bank estimate).

Figure 1. Current health expenditure (per capita) and mobile health information seeking in 10 Asian countries. The size of the nodes represents the Information Communications Technology Development Index (IDI), and the shading of the nodes represents the Gini index of the countries.



Discussion

Principal Findings

Understanding how individual- and country-level differences in MHIS are affected by socioeconomic conditions is important for developing and evaluating public health policy. Studies have suggested that the digital divide is a serious barrier that largely contributes to health inequality [42]. In our study, we extended these findings by examining MHIS in Asia and the digital divide associated with such behavior at both the individual and country levels. Based on a sample of 9068 participants from 10 Asian countries, our results suggest that the act of seeking health information with smartphones varies according to individuals' SES, perceptions of technology, and country of residence. It seems that smartphone technology widens the digital divide throughout the socioeconomic structure of society, such that individuals who are of higher education or subjective SES, married, parents, and/or employed are more likely to use smartphones to seek health information. However, smartphones also bridge the gap between countries to some extent. Individuals from countries with lower expenditures in health are more likely to use smartphones to seek health information. The multilevel digital divide documented in our study has practical implications for public health professionals.

Profiling Mobile Health Information Seekers in Asia

Being a woman has consistently predicted increased activity in seeking health information online [13-16]. Women tend to use smartphones for obtaining health information more often than men owing to their higher engagement in health-related activities for themselves and their family members. In that light, our study extends past findings to the Asian context. Interestingly, although age has been emphasized as an important factor of social division in previous studies [6], its effect was not significant in our regression models, possibly due to the nonlinear relationship between age and MHIS. Younger and older generations emerged as being more likely to use smartphones to seek health information than middle-aged individuals. This may reflect the fact that the younger generation is more familiar and comfortable with using smartphones to meet their everyday needs, including health-related needs, whereas the older generation has a strong motivation and need to seek health information from any source.

Our findings also highlight the important role of technological factors in explaining MHIS. The perceived innovativeness of technology was a primary factor for accepting new technology, including in relation to mHealth. Information quality surfaced as another factor related to seeking health information, because the information involved in these practices is highly personal and thus sensitive. However, previous studies on mHealth have not sufficiently investigated the role of these technological factors [8]. Future studies aiming for a comprehensive understanding of MHIS should thus include the factors of technology innovativeness and information quality.

Individual-Level Digital Divide

Our study confirmed that individuals who are of higher education or social status, married, parents, and/or employed are more likely to use smartphones to seek health information. Therefore, our results extend previous findings to the Asian context. The digital divide in health refers to inequalities not only in internet access, mobile technology, and social media but also in the ability to comprehend the information found online. Although access to the internet or smartphones is now ubiquitous in most Asian countries, a second level of socioeconomic inequalities such as different levels of education and SES affect individuals' ability to seek and comprehend online information [8,28]. Therefore, future studies should focus on this second level of the digital divide and its influence on mHealth across various social groups. Public health efforts attempting to leverage the power of mobile technology should also adopt different strategies to avoid inequalities across social structures. For instance, online communication–based interventions should better investigate and address issues

pertaining to eHealth literacy, including the ability to seek, find, understand, and appraise health information from electronic sources, so as to reduce inequalities in communication across different socioeconomic groups.

Country-Level Digital Divide

At the country level, we found that participants from countries that spend less on health per capita were more likely to rely on smartphones as a source for health information. This result stresses that smartphones in developing countries, including China, India, the Philippines, Indonesia, Vietnam, and Thailand, may function as a tool for managing daily activities, including seeking health information. Considering that these Asian countries have a relatively high mobile internet penetration rate [43], our results provide evidence that smartphones act as tools that can bridge health inequalities between countries. In developing countries where health or digital resources are limited, mobile technology may also help individuals to access information about health as well as manage their health. Whereas previous studies investigating the digital divide in access to health information and technology have focused on the socioeconomic characteristics, internet access, and information literacy of individuals, our study examined the digital divide at both the individual and country levels. Therefore, this study adds a new dimension for understanding the digital divide in mHealth.

Strengths and Limitations

The strengths of our study include its large sample, and the collection of data on individuals' smartphone use and perceptions in several Asian countries. A previous meta-analysis suggested that studies about information-seeking behavior with mobile technology in Asia with large samples have been lacking [8]. Thus, our study has filled this gap through administering by far the largest multinational survey on smartphone users in Asia. Another strength was that two levels of the digital divide were examined, which furnishes considerable knowledge about the digital divide in MHIS.

Nevertheless, the limitations of our study should be noted when generalizing the findings. First, our sample was not representative. Our participants were smartphone owners who live in urban areas in each country sampled. Despite variance in our participants' SES, their relatively high SES may limit the generalizability of the findings. Second, we only examined the role of socioeconomic factors and technology perceptions on MHIS. Other psychological-related factors (eg, attitude, self-efficacy, perceived risk, worry, and anxiety) should also be taken into consideration in future studies with the aim of forming a comprehensive understanding of such phenomena [8]. Finally, information quality has drawn great attention from both academia and industry because of prevalent misinformation online [44]. Although our study found a marginal effect of information quality, future studies aiming to gain a nuanced understanding of such a phenomenon could examine the multidimensional nature of information quality or the reciprocal relationship between perceived online information quality and information seeking.

Conclusion

Among smartphone users in Asia, seeking health information on mobile devices varies according to the users' SES, perceptions of technology and information, and their governments' health expenditures, but not in accordance with the ICT divide or economic inequality at the country level. These findings suggest that although smartphones represent a readily available source of health information, they can also create inequalities in the access to health information among different socioeconomic classes of society. At the same time, the findings imply that smartphones are widely accepted as a tool for daily activities and communication in developing areas in Asia. In that light, mobile technology appears to be a particularly useful complement for the management of health in developing countries.

Acknowledgments

The study was supported by the Institute on Asian Consumer Insight funded jointly by the Singapore Economic Development Board and Nanyang Technological University.

Conflicts of Interest

None declared.

References

1. Watts S, Mackenzie A, Thomas C, Griskaitis A, Mewton L, Williams A, et al. CBT for depression: a pilot RCT comparing mobile phone vs. computer. BMC Psychiatry 2013 Feb 07;13:49 [FREE Full text] [doi: 10.1186/1471-244X-13-49] [Medline: 23391304]
2. Brown K, Campbell SW, Ling R. Mobile phones bridging the digital divide for teens in the US? Future Internet 2011 May 13;3(2):144-158. [doi: 10.3390/fi3020144]
3. Xia Q, Sifah EB, Asamoah KO, Gao J, Du X, Guizani M. MeDShare: trust-less medical data sharing among cloud service providers via blockchain. IEEE Access 2017;5:14757-14767. [doi: 10.1109/access.2017.2730843]
4. Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. Info Systems J 2018 Jan 29;28(6):995-1019. [doi: 10.1111/isj.12179]

5. Feng Y, Xie W. Digital divide 2.0: the role of social networking sites in seeking health information online from a longitudinal perspective. J Health Commun 2015 Aug 13;20(1):60-68. [doi: 10.1080/10810730.2014.906522] [Medline: 25119019]
6. Kontos E, Blake KD, Chou WYS, Prestin A. Predictors of eHealth usage: insights on the digital divide from the Health Information National Trends Survey 2012. J Med Internet Res 2014 Jul 16;16(7):e172 [FREE Full text] [doi: 10.2196/jmir.3117] [Medline: 25048379]
7. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
8. Wang X, Shi J, Kong H. Online health information seeking: a review and meta-analysis. Health Commun 2021 Sep;36(10):1163-1175. [doi: 10.1080/10410236.2020.1748829] [Medline: 32290679]
9. Pian W, Song S, Zhang Y. Consumer health information needs: A systematic review of measures. Inf Process Manag 2020 Mar;57(2):102077. [doi: 10.1016/j.ipm.2019.102077]
10. Chi J, Pian W, Zhang S. Consumer health information needs: a systematic review of instrument development. Inf Process Manag 2020 Nov;57(6):102376. [doi: 10.1016/j.ipm.2020.102376]
11. Zhao Y, Ni Q, Zhou R. What factors influence the mobile health service adoption? A meta-analysis and the moderating role of age. Int J Inf Manag 2018 Dec;43:342-350. [doi: 10.1016/j.ijinfomgt.2017.08.006]
12. Wang X, Lee KM. The paradox of technology innovativeness and risk perceptions – A profile of Asian smartphone users. Telemat Inform 2020 Aug;51:101415. [doi: 10.1016/j.tele.2020.101415]
13. Sun Y, Wang N, Guo X, Peng Z. Understanding the acceptance of mobile health services: a comparison and integration of alternative models. J Electron Commer Res 2013;14:183-200.
14. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
15. Chae J. A comprehensive profile of those who have health-related apps. Health Educ Behav 2018 Aug 20;45(4):591-598. [doi: 10.1177/1090198117752784] [Medline: 29353544]
16. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
17. Deng Z, Liu S. Understanding consumer health information-seeking behavior from the perspective of the risk perception attitude framework and social support in mobile social media websites. Int J Med Inform 2017 Sep;105:98-109. [doi: 10.1016/j.ijmedinf.2017.05.014] [Medline: 28750916]
18. Dwivedi YK, Shareef MA, Simintiras AC, Lal B, Weerakkody V. A generalised adoption model for services: A cross-country comparison of mobile health (m-health). Gov Inf Quart 2016 Jan;33(1):174-187. [doi: 10.1016/j.giq.2015.06.003]
19. Guo X, Han X, Zhang X, Dang Y, Chen C. Investigating m-Health acceptance from a protection motivation theory perspective: gender and age differences. Telemed J E Health 2015 Aug;21(8):661-669. [doi: 10.1089/tmj.2014.0166] [Medline: 25919800]
20. Lee YW, Strong DM, Kahn BK, Wang RY. AIMQ: a methodology for information quality assessment. Inf Manage 2002 Dec;40(2):133-146. [doi: 10.1016/s0378-7206(02)00043-5]
21. DeLone WH, McLean ER. Information systems success: the quest for the dependent variable. Inf Syst Res 1992 Mar;3(1):60-95. [doi: 10.1287/isre.3.1.60]
22. Miller LMS, Bell RA. Online health information seeking: the influence of age, information trustworthiness, and search challenges. J Aging Health 2012 Apr;24(3):525-541. [doi: 10.1177/0898264311428167] [Medline: 22187092]
23. Agarwal R, Prasad J. A conceptual and operational definition of personal innovativeness in the domain of information technology. Inf Syst Res 1998 Jun;9(2):204-215. [doi: 10.1287/isre.9.2.204]
24. Xu H, Luo XR, Carroll JM, Rosson MB. The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. Decision Support Syst 2011 Apr;51(1):42-52. [doi: 10.1016/j.dss.2010.11.017]
25. Rai A, Chen L, Pye J, Baird A. Understanding determinants of consumer mobile health usage intentions, assimilation, and channel preferences. J Med Internet Res 2013 Aug 02;15(8):e149 [FREE Full text] [doi: 10.2196/jmir.2635] [Medline: 23912839]
26. Renahy E, Parizot I, Chauvin P. Health information seeking on the internet: a double divide? Results from a representative survey in the Paris metropolitan area, France, 2005-2006. BMC Public Health 2008 Feb 21;8:69 [FREE Full text] [doi: 10.1186/1471-2458-8-69] [Medline: 18291024]
27. Scheerder A, van Deursen A, van Dijk J. Determinants of internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. Telemat Inform 2017 Dec;34(8):1607-1624. [doi: 10.1016/j.tele.2017.07.007]
28. Bodie GD, Dutta MJ. Understanding health literacy for strategic health marketing: eHealth literacy, health disparities, and the digital divide. Health Mark Q 2008;25(1-2):175-203. [doi: 10.1080/07359680802126301] [Medline: 18935884]
29. Bert F, Giacometti M, Gualano MR, Siliquini R. Smartphones and health promotion: a review of the evidence. J Med Syst 2014 Jan;38(1):9995. [doi: 10.1007/s10916-013-9995-7] [Medline: 24346929]
30. Graetz I, Huang J, Brand RJ, Hsu J, Yamin CK, Reed ME. Bridging the digital divide: mobile access to personal health records among patients with diabetes. Am J Manag Care 2018 Jan;24(1):43-48 [FREE Full text] [Medline: 29350505]
31. Tsetsi E, Rains SA. Smartphone internet access and use: extending the digital divide and usage gap. Mobile Media Commun 2017 Jun 13;5(3):239-255. [doi: 10.1177/2050157917708329]

32. Free C, Phillips G, Felix L, Galli L, Patel V, Edwards P. The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. BMC Res Notes 2010 Oct 06;3:250 [FREE Full text] [doi: 10.1186/1756-0500-3-250] [Medline: 20925916]
33. Ember C, Ember M. Cross-cultural research. In: Bernard HR, editor. Handbook of methods in cultural anthropology. Walnut Creek, CA: AltaMira Press; 1998:647-687.
34. Lopez-Fernandez O, Kuss DJ, Romo L, Morvan Y, Kern L, Graziani P, et al. Self-reported dependence on mobile phones in young adults: A European cross-cultural empirical survey. J Behav Addict 2017 Jun 01;6(2):168-177 [FREE Full text] [doi: 10.1556/2006.6.2017.020] [Medline: 28425777]
35. Doshi T, Smalls BL, Williams JS, Wolfman TE, Egede LE. Socioeconomic status and cardiovascular risk control in adults with diabetes. Am J Med Sci 2016 Jul;352(1):36-44 [FREE Full text] [doi: 10.1016/j.amjms.2016.03.020] [Medline: 27432033]
36. Operario D, Adler NE, Williams DR. Subjective social status: reliability and predictive utility for global health. Psychol Health 2007 Feb;19(2):237-246. [doi: 10.1080/08870440310001638098]
37. Limayem M, Khalifa M, Frini A. What makes consumers buy from internet? A longitudinal study of online shopping. IEEE Trans Syst Man Cybern A 2017;30(4):421-432. [doi: 10.1109/3468.852436]
38. The future of truth and misinformation online. Pew Research Center. 2017. URL: http://www.pewinternet.org/2017/10/19/the-future-of-truth-and-misinformation-online/ [accessed 2020-10-24]
39. ICT Development Index 2017. International Telecommunication Union (ITU). URL: https://www.itu.int/net4/ITU-D/idi/ [accessed 2021-12-30]
40. Indicators. The World Bank Data. URL: https://data.worldbank.org/indicator [accessed 2021-12-30]
41. Global health expenditure database. World Health Organization. URL: https://apps.who.int/nha/database [accessed 2021-12-30]
42. Latulippe K, Hamel C, Giroux D. Social health inequalities and eHealth: a literature review with qualitative synthesis of theoretical and empirical studies. J Med Internet Res 2017 Apr 27;19(4):e136 [FREE Full text] [doi: 10.2196/jmir.6731] [Medline: 28450271]
43. Newzoo Global Mobile Market Report. Newzoo. 2019. URL: https://newzoo.com/insights/trend-reports/newzoo-global-mobile-market-report-2019-light-version/ [accessed 2021-10-24]
44. Chen L, Wang X, Peng T. Nature and diffusion of gynecologic cancer-related misinformation on social media: analysis of tweets. J Med Internet Res 2018 Oct 16;20(10):e11515 [FREE Full text] [doi: 10.2196/11515] [Medline: 30327289]

Abbreviations

CHE: current health expenditure
ICT: information communication technology
IDI: Information Communication Technology Development Index
mHealth: mobile health
MHIS: mobile health information seeking
SES: socioeconomic status

Edited by R Kukafka; submitted 03.09.20; peer-reviewed by W Pian, H Kim; comments to author 02.11.20; revised version received 31.12.20; accepted 21.06.21; published 13.01.22.

Please cite as:
Wang X, Shi J, Lee KM
The Digital Divide and Seeking Health Information on Smartphones in Asia: Survey Study of Ten Countries
J Med Internet Res 2022;24(1):e24086
URL: https://www.jmir.org/2022/1/e24086
doi:10.2196/24086
PMID:35023845

Original Paper

Impact of Internet Use on Cognitive Decline in Middle-Aged and Older Adults in China: Longitudinal Observational Study

Xinyue Yu[1], MSc; Aruhan Mu[1], MRES; Xiang Wu[1], PhD; Liqin Zhou[1], PhD

School of Medicine and Health Management, Huazhong University of Science and Technology, Wuhan, China

Corresponding Author:
Liqin Zhou, PhD
School of Medicine and Health Management
Huazhong University of Science and Technology
13 Hangkong Road
Wuhan, 430030
China
Phone: 86 15927686325
Email: zhoulq@hust.edu.cn

Abstract

Background: Given that cognitive decline lacks effective treatment options and has severe implications for healthy aging, internet use may achieve nonpharmacological relief of cognitive decline through cognitive stimulation and social engagement.

Objective: This longitudinal study aimed to investigate the relationship between the diversity, frequency, and type of internet use and cognitive decline, and to provide theoretical support and suggestions for mitigating cognitive decline in middle-aged and older adults.

Methods: Data were obtained from a total of 10,532 survey respondents from the China Family Panel Studies database from wave 3 (2014) and wave 5 (2018) of the survey. Cognitive function was measured using vocabulary tests, and internet use was categorized into five aspects: study, work, socializing, entertainment, and commercial-related activities. Associations between the diversity, frequency, and type of internet use and cognitive decline were estimated by controlling for demographic variables and health status risk factors through fixed-effects models.

Results: After controlling for demographic and health status risk factors, the type and frequency of internet use were found to be associated with cognitive functioning during the subsequent 4-year period, and different types of internet use had different effects on cognitive decline. Frequency of internet use of at least once a week for study (β=0.620, 95% CI 0.061 to 1.180; *P*=.04), work (β=0.896, 95% CI 0.271 to 1.520; *P*=.01), and entertainment (β=0.385, 95% CI –0.008 to 0.778; *P*=.06), as well as less than once a week for social purposes (β=0.860, 95% CI 0.074 to 1.650; *P*=.06), were associated with better cognitive function. Frequency of internet use of less than once a week for commercial-related activities (β=–0.906, 95% CI –1.480 to –0.337; *P*=.005) was associated with poorer cognitive function. Using the internet for more than one type of activity (β=0.458, 95% CI 0.065 to 0.850; *P*=.03) and at least once a week (β=0.436, 95% CI 0.066 to 0.806; *P*=.02) was associated with better cognitive function.

Conclusions: This study shows that breadth and depth of internet use are positively associated with cognitive function and that different types of internet use have different roles in cognitive decline. The importance of the internet as a nonpharmacological intervention pathway for cognitive decline is emphasized. Future research could explore specific mechanisms of influence.

(J Med Internet Res 2022;24(1):e25760) doi:10.2196/25760

KEYWORDS

internet use; cognitive decline; China; fixed-effects analysis

Introduction

Cognitive decline is an irreversible process of pathological changes in the brain that often starts in individuals aged 45 to 60 years [1]. The aggravation of cognitive decline is likely to lead to dementia as well as physical disability and death [2]. Some studies have shown that almost 20% of Chinese people over 60 years of age have mild cognitive decline and develop dementia at a rate of 6% every year [3]. Although there is no effective treatment for cognitive decline, neuroscience and cognitive aging studies have shown that patients with cognitive decline retain some cognitive abilities and plasticity [4,5]. The results of an intervention evaluation on the impact of global

dementia interventions showed that interventions that delay disease onset and progression by 1 year would reduce the incidence of dementia by 9.2 million cases in 2025 [6]. Cognitive decline is not only a significant burden on society and the economy but also carries the pressure of informal care costs and health resources. Therefore, the use of nonpharmacological interventions for middle-aged and older adults aged 45 years and above is of great significance in reducing the occurrence of cognitive decline, slowing down cognitive decline, and even reversing the disease [7].

In studies exploring risk factors for cognitive decline, demographic variables, individual health status, and social factors are largely taken into account, and the conclusions reached are largely consistent. Among them, demographic variables, such as age and education, are risk factors for cognitive decline. Physical diseases, such as cardiovascular disease and stroke [8]; unhealthy behaviors, such as smoking and alcohol consumption [9-12]; and social isolation, loneliness, depression, and other conditions [13-16] can also harm cognitive function in middle-aged and older adults. The World Health Organization (WHO), in its Guidelines for Mitigating the Risk of Cognitive Decline and Dementia, recommends reducing the risk of cognitive decline by living a healthy lifestyle. However, there are individual differences among older people, including reduced mobility, physical deterioration, social isolation, and differences in their surroundings, resulting in limited access and low availability of resources [17]. The increase in internet penetration has allowed internet use to gradually penetrate middle and older age groups [18]. The researchers are interested in the impact of internet use as a cognitive stimulus on middle-aged and older adults, providing an opportunity to mitigate cognitive decline. This measure is highly actionable.

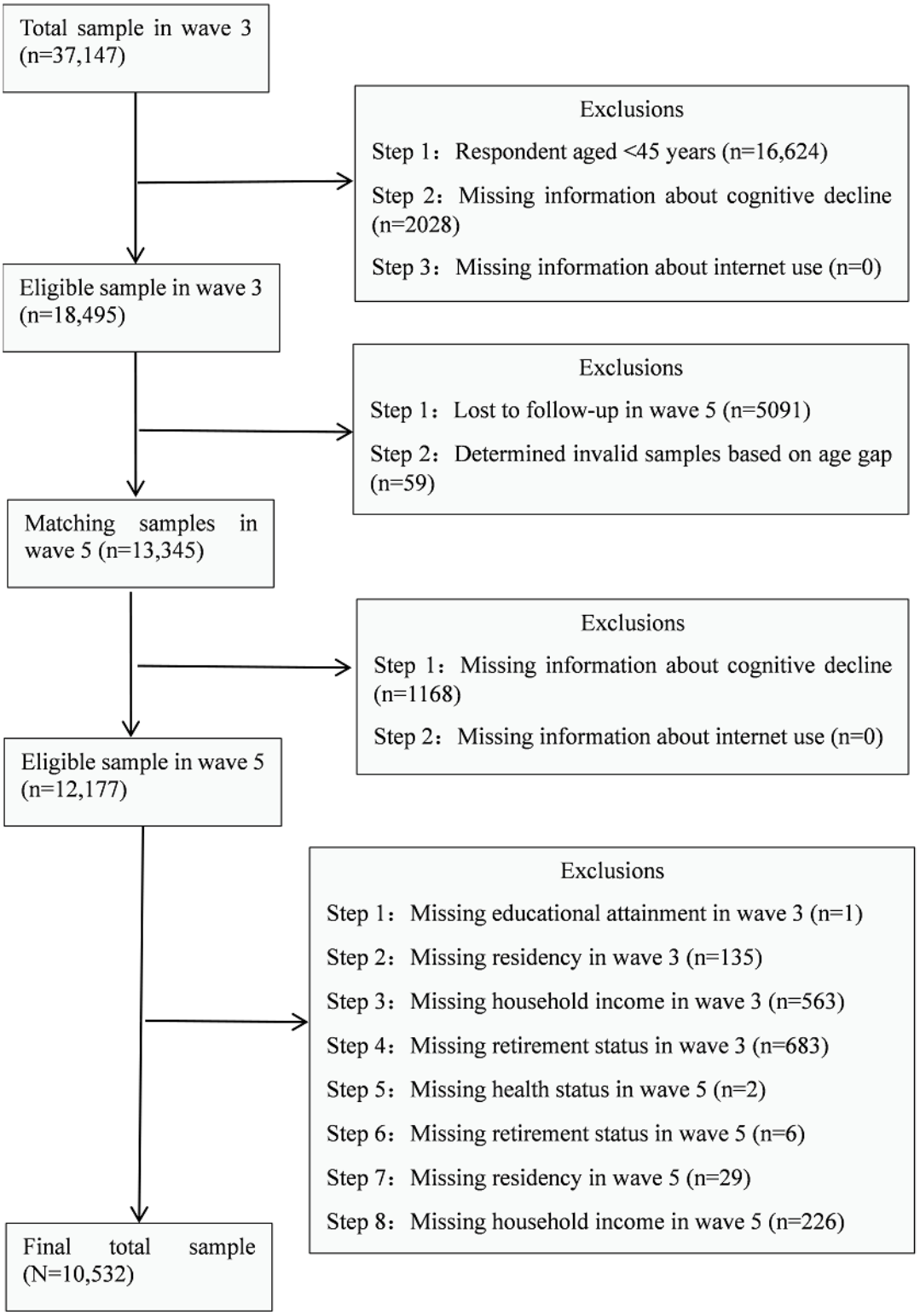
Over the past few years, researchers have conducted several studies on the effects of internet use on cognition. Research shows that using the internet can improve the choice ability of older adults with cognitive decline. Training older adults to use computers and the internet can positively impact their quality of life [18]. Owning electronic devices can reduce cognitive decline in older adults [17,19]; older adults' use of email, serious video games, and virtual reality may mitigate cognitive decline [19,20]. Learning and using online social networking sites can be used as interventions to maintain or enhance cognitive function in older adults [21]. The use of social networking sites by older adults will improve perceived social support and connection, and reduce perceived social isolation [22]. Previous studies mainly measured internet use as the use of email or digital devices. They used cross-sectional data more often to determine the impact of internet use on cognitive decline. In the Chinese context, email is hardly representative of internet use. China Family Panel Studies (CFPS) is a biennial prospective observational study; this study used 2014 data as the baseline and 2018 data as the follow-up to investigate the following: (1) whether five types of internet use were independently associated with cognitive decline in middle-aged and older adults and (2) whether frequency and diversity of internet use were independently associated with cognitive decline in middle-aged and older adults.

Methods

Sample and Data Collection

The data for this study were from the CFPS database survey data that were collected in 2014 (wave 3) and 2018 (wave 5) [23]. The 2 years of survey data were collected using the same cognitive test questions, and both tested the internet use module. The CFPS baseline (2010) sample covers 25 provinces, municipalities, or autonomous regions, representing 95% of China's population. The researchers who conducted the 2010 baseline survey interviewed a total of 14,960 households and 42,590 individuals, and they launched a long-term follow-up survey of individual samples. Data cleaning resulted in a sample of 37,147 respondents. This study consisted of 10,532 adults aged 45 years or older in 2014. We used data from wave 3 in 2014 as baseline data, and we used data from wave 5 in 2018 as follow-up data, which had been collected after a 4-year follow-up period. Cognitive function was tested using the same questions at baseline testing and during follow-up. We further excluded samples according to the following criteria: (1) respondents aged less than 45 years at baseline (n=16,624), (2) no information provided on cognitive function (n=3196) or internet use (n=0), (3) loss of respondents after matching the two waves (n=5091), and (4) missing sample of control variables (n=1704). A total of 10,532 respondents were finally screened. Figure 1 illustrates the exclusion criteria and the process of respondent screening.

Figure 1. Flowchart showing respondent screening.



Measures

Dependent Variable

Cognitive function was the dependent variable. In order to assess the level of cognitive function, we used a vocabulary test. The test, which consisted of 34 Chinese characters, sought to measure one's vocabulary by testing a respondent's ability to recognize difficult characters; the score ranged from 0 to 34 [24].

Independent Variables: Internet Use

In this study, we examined internet use from the following three aspects:

1. Type of internet use. This item included the frequency of using the internet for the following five topics: study, work, socializing, entertainment, and commercial-related activities. For example, "In general, how frequently do you use the internet to study, such as searching for study

materials or online study courses?" The survey participants selected a frequency to reflect their use of the internet for each purpose. The variable was coded as 0 (none), 1 (less than once a week), or 2 (at least once a week).

2. Diversity of internet use types. This represented the participant's total number of different internet use types, which were coded as 0 (none), 1 (one type), or 2 (more than one type).
3. Frequency of internet use. This represented the participant's maximum frequency of internet use. We merged the groups who reported a frequency of more than once a week and recoded the responses as "at least once a week." Groups who reported a frequency of less than once a week were combined, and the responses were recoded as "less than once a week." Therefore, the variable was coded as 0 (none), 1 (less than once a week), or 2 (at least once a week).

Control Variables

A set of variables previously revealed to be associated with internet use and cognitive function (ie, gender, age, educational attainment, marital status, household income, residency, health status, and retirement status) were controlled to provide more compelling evidence by reducing the amount of confounding influence they might have on the effects of internet use. Gender, marital status, residency, and retirement status were defined as dichotomous variables (gender: 0 [man] or 1 [woman]; marital status: 0 [single] or 1 [married or living as married]; residency: 0 [rural] or 1 [urban]; retirement status: 0 [no] or 1 [yes]). Age was measured in years, and educational attainment was measured by asking the respondent to indicate the highest level of education completed: 1 (illiterate), 2 (primary school), 3 (junior high school), 4 (senior high school), or 5 (college and above). Health status was measured according to respondents' self-rated physical health status: 1 (poor), 2 (fair), 3 (good), 4 (very good), or 5 (excellent). Household income was the total income from all types of work and nonwork for all household members divided by the number of household members; this was log-transformed because of its highly skewed distribution.

Data Analysis

Descriptive analyses were conducted to summarize the demographic characteristics of the study sample. Participants' characteristics were summarized using mean (SD) for continuous variables and proportions for categorical variables. We aimed to determine whether there was a significant difference in the highest frequency and diversity of internet use between different groups. In this case, the chi-square test was used for categorical variables, and we used the Kruskal-Wallis test for all continuous variables.

We used the Hausmann test, which is more suitable for panel data processing, to implement the fixed-effects model, taking into account both individual and time effects of the sample. The measurement model is constructed as follows:

$$CognitionFunction_{it} = \beta_1 Internet_{it} + \beta_2 x_{it} + \alpha_i + \mu_t + \varepsilon_{it}$$

$CognitionFunction_{it}$ represents the score of individual i on a word test of cognitive function at time t; $Internet_{it}$ represents individual i's purpose, frequency, and diversity of different internet uses at time t; x_{it} represents characteristics of individual i over time, including marital status, retirement status, income, and general health; α_i represents characteristics of individual i that do not change over time, including gender, education level, and possible unobservable effects; μ_t represents the effect that time t does not vary from individual to individual; and ε_{it} represents the error term.

The main model consisted of three regression equations measuring the following three relationships: the relationship between the diversity of respondents' internet use and their cognitive function, the relationship between respondents' maximum frequency and their cognitive function, and the relationship between respondents' frequency of internet use for different purposes and their cognitive function. Control variables were included throughout the analyses. All analyses were performed using R (version 4.0.2; The R Foundation).

Results

Descriptive Statistics

The basic characteristics of the whole sample population (N=10,532) at baseline for this study (ie, wave 3) are shown in Table 1. The mean age of the respondents was 58.7 (SD 9.3) years, 5405 (51.3%) participants were female, and 3526 (33.5%) participants were retired. A total of 10,202 (96.9%) participants had no college education, 9518 (90.4%) were married or cohabiting, and 6456 (61.3%) considered their health to be good. The mean household income of the participants was ¥12,737.30 (SD ¥16,569.10; mean US $1999.39, SD US $2600.87). Moreover, we found that all covariates were significantly associated with the frequency of internet use. In addition, we observed that all covariates were significantly associated with the frequency of respondents' internet use by chi-square test and Kruskal-Wallis test. This is consistent with the results of previous studies.

Table 1. Sample characteristics of respondents at baseline.

Variable	Value (N=10,532)	Frequency of internet use		Diversity of internet use	
		Chi-square[a] (*df*) or Kruskal-Wallis[b] (*df*)	*P* value	Chi-square[a] (*df*) or Kruskal-Wallis[b] (*df*)	*P* value
Gender, n (%)					
Male	5127 (48.7)	39.7[c] (2)	<.001[c]	38.9 (2)	<.001
Female	5405 (51.3)				
Age (years), mean (SD)	58.7 (9.3)	393.0 (2)	<.001	406.0 (2)	<.001
Educational attainment, n (%)					
Illiterate	4001 (38.0)	2077.7 (8)	<.001	2227.0 (8)	<.001
Primary school	2459 (23.4)				
Junior high school	2576 (24.5)				
Senior high school	1166 (11.1)				
College and above	330 (3.1)				
Marital status, n (%)					
Single	1014 (9.6)	10.4 (2)	.006	10.7 (2)	.005
Married or living as married	9518 (90.4)				
Residency, n (%)					
Rural	5916 (56.2)	508.7 (2)	<.001	508.4 (2)	<.001
Urban	4616 (43.8)				
Retirement status, n (%)					
No	7006 (66.5)	33.6 (2)	<.001	43.8 (2)	<.001
Yes	3526 (33.5)				
Household income (¥[d]), mean (SD)	12,737.30 (16,569.10)	547.0 (2)	<.001	547.0 (2)	<.001
Health status, n (%)					
Poor	2297 (21.8)	126.1 (8)	<.001	128.3 (8)	<.001
Fair	1179 (16.9)				
Good	3760 (35.7)				
Very good	1630 (15.5)				
Excellent	1066 (10.1)				

[a]Chi-square tests were performed for gender, educational attainment, marital status, residency, retirement status, and health status.

[b]Kruskal-Wallis tests were performed for age and household income.

[c]The statistical test value and the *P* value for a group of variables is listed in the top row of that group.

[d]A currency exchange rate of US $1=¥6.37 is applicable.

Internet Use Among Older Adults

Table 2 shows data from two waves regarding the diversity, frequency, and type of internet use among the middle-aged and older adult population (N=10,532). The participation rates show that respondents did not use the internet to a very high degree, while in 2018, the usage rates saw a slight increase of 14.9%. However, there were still more than 75% of middle-aged and older adults who did not use the internet. In 2014, the lowest rate of internet use was for commercial-related activities (n=245, 2.3%), and the highest rate of internet use was for entertainment (n=694, 6.6%). In 2018, the lowest rate of internet use was for work (n=620, 5.9%), and the highest rate of internet use was for socializing (n=2131, 20.2%). The 2-year comparison shows that the middle-aged and older adult respondents who used the internet for socializing had the highest growth rate of 15.3%, while those who used the internet for work had the lowest growth rate of 2.3%.

Table 2. Internet use type, frequency, and diversity by respondents for each of the two waves.

Internet use	Participants (N=10,532), n (%)	
	Wave 3 (2014)	Wave 5 (2018)
Type		
Study		
None	9959 (94.6)	9609 (91.2)
Less than once a week	104 (1.0)	220 (2.1)
At least once a week	469 (4.4)	703 (6.7)
Work		
None	10,149 (96.4)	9912 (94.1)
Less than once a week	70 (0.7)	84 (0.8)
At least once a week	313 (3.0)	536 (5.1)
Socializing		
None	10,012 (95.1)	8401 (79.8)
Less than once a week	102 (1.0)	182 (1.7)
At least once a week	418 (4.0)	1949 (18.5)
Entertainment		
None	9838 (93.4)	8449 (80.2)
Less than once a week	160 (1.5)	223 (2.1)
At least once a week	534 (5.1)	1860 (17.7)
Commercial-related activities		
None	10,287 (98.2)	9631 (91.5)
Less than once a week	191 (1.8)	507 (4.8)
At least once a week	54 (0.5)	394 (3.7)
Diversity		
None	9651 (91.6)	8081 (76.7)
One type	176 (1.7)	401 (3.8)
More than one type	705 (6.7)	2050 (19.5)
Frequency		
None	9651 (91.6)	8081 (76.7)
Less than once a week	73 (0.7)	105 (1.0)
At least once a week	808 (7.7)	2346 (22.3)

In 2018, as compared to 2014, there was a 2.1% increase in the number of subjects participating in one type of internet activity and a 12.8% increase in the number of subjects participating in more than one kind of online activity. The participation of people 45 years of age and older in internet activities increased in breadth. The subjects' frequent participation (ie, at least once a week) in online activities increased by 14.6% in 2018, as compared to 2014, while participation of less than once a week increased by 0.3%. Internet activity participation by people aged 45 years and over gradually shifted to deeper involvement.

Relationship Between Internet Use and Cognitive Function

Table 3 outlines the contemporary association between cognitive function and the diversity, frequency, and type of internet use between waves 3 and 5 with other time-varying confounders controlled. We included covariates in a fixed-effects model in which residency and health status emerged as significantly associated with cognitive function. Urban users were associated with better cognitive function compared to rural users (β=0.687, *P*=.03). Respondents in better health had better cognitive functioning compared to those in poor health, with a health status rating of "excellent" being most associated with better cognitive functioning (β=0.602, *P*=.01). Our variables were all categorical variables, and the rating "none" represents the baseline level. The results of the fixed-effects model showed a relationship between cognitive function and frequencies of internet use of "less than once a week" and "at least once a week," as compared to a frequency of "none." Participating in multiple types of internet activities and engaging in internet

activities at least once a week were associated with better cognitive function in middle-aged and older adults (β=0.458, *P*=.03; β=0.436, *P*=.02), whereas participating in a single type of internet activity and using the internet less than once a week were not associated with cognitive function. Using the internet at least once a week for study and work was associated with better cognitive function (β=0.620, *P*=.04; β=0.896, *P*=.01), whereas using the internet less than once a week for study failed to suggest a relationship and using the internet less than once a week for work still predicted better cognitive function (β=0.955, *P*=.11). Using the internet less than once a week for socializing and using it at least once a week for entertainment were associated with better cognitive function (β=0.860, *P*=.06; β=0.385, *P*=.06), but no relationship could be observed between better cognitive function and using the internet at least once a week for socializing and less than once a week for entertainment. Using the internet less than once a week for commercial-related activities was associated with worse cognitive function (β=–0.906, *P*=.005). No relationship could be observed between using the internet more than once a week for commercial-related activities and better cognitive function.

Table 3. Associations between cognitive function and internet use diversity, frequency, and type using fixed-effects regression.

Independent variables: internet use[a,b]	Dependent variable: cognitive function	
	β (95% CI)	*P* value
Diversity		
None	Reference	N/A[c]
One type	0.319 (–0.328 to 0.966)	.33
More than one type	0.458 (0.065 to 0.850)	.03
Frequency		
None	Reference	N/A
Less than once a week	0.267 (–0.893 to 1.430)	.64
At least once a week	0.436 (0.066 to 0.806)	.02
Type		
Study		
None	Reference	N/A
Less than once a week	0.529 (–0.202 to 1.260)	.23
At least once a week	0.620 (0.061 to 1.180)	.04
Work		
None	Reference	N/A
Less than once a week	0.955 (–0.087 to 2.000)	.11
At least once a week	0.896 (0.271 to 1.520)	.01
Socializing		
None	Reference	N/A
Less than once a week	0.860 (0.074 to 1.650)	.06
At least once a week	0.004 (–0.364 to 0.372)	.98
Entertainment		
None	Reference	N/A
Less than once a week	0.512 (–0.242 to 1.270)	.21
At least once a week	0.385 (–0.008 to 0.778)	.06
Commercial-related activities		
None	Reference	N/A
Less than once a week	–0.906 (–1.480 to –0.337)	.005
At least once a week	–0.566 (–1.280 to 0.152)	.15

[a]All models controlled for time-varying variables, including wave, gender, educational attainment, marital status, household income, residency, health status, and retirement status.

[b]There were 21,064 observations, and R^2=0.002.

[c]N/A: not applicable; this row is the reference value.

Discussion

Principal Findings

This study was nationally representative and showed longitudinal protective associations between different internet use dimensions and cognitive decline among the older adult population in China. The global population is aging rapidly, and China is a country with a large older adult population, of which 177 million are over 65 years of age. According to WHO estimates, the proportion of the world's population over 60 years of age will double between 2000 and 2050, from 11% to 22% [7]. Therefore, as the aging population is growing, the use of nonpharmacological interventions for middle-aged and older adults aged 45 years and above is of great significance in reducing the incidence of cognitive decline, delaying cognitive decline, and reversing the disease in middle-aged and older adults. As cognitive decline incidence continues to rise, the morbidity age becomes lower and the medical burden continues to increase. What is more, the disease is more likely to develop into dementia without effective drug treatment. As of June 2019, the number of internet users in China reached 854 million. The

proportion of internet users aged 50 years and above increased from 12.5% to 13.6% at the end of 2018, and internet penetration continued among middle-aged and older adult age groups. Internet use offers opportunities and high accessibility to mitigate cognitive decline among older adults. Based on a nationally representative survey by CFPS, we found that the correlation between internet use and cognitive decline is still present in China. The different effects of the internet use dimensions on cognitive decline are worth discussing.

We found that internet users had better cognitive function compared to nonusers at the 4-year follow-up. Evidence in other settings includes using five periods of data from the English Longitudinal Study of Aging (ELSA) database, with cognitive function as the dependent variable and whether they used the internet or email as the independent variable, to conclude that internet use helped reduce the number of people aged 50 to 89 years with cognitive decline [20]. Although there is no definitive mechanism that affects cognitive decline, there are two widely accepted mechanisms in the research to explain this. The first is the cognitive reserve hypothesis [25-27], which views internet use as a cognitively stimulating activity that uses the brain's neural networks to decrease brain damage. The second is the stress hypothesis [28], which views the internet as a form of social engagement, where the sense of belonging and constructed social networks generated by the activity stimulate brain evolution and functional development. Failure to adapt to stress, which is associated with the pathogenesis of dementia, increases an individual's glucocorticoid levels; this increase in glucocorticoids leads to hippocampal damage, which can impair an individual's cognitive function [29]. These hypotheses explain, to some extent, the results of this study.

We built upon previous research to verify that different residency and health statuses in middle-aged and older adults are associated with different cognitive functions. Older adults who live in urban settings and are in good health are less likely to experience cognitive decline. Diversity and frequency of internet use have differential relationships on cognitive decline. More extensive and frequent use of the internet as a cognitively stimulating activity is associated with better mental functioning.

The use of the internet for study and work shows that it can be a tool for accessing and processing information. We have observed that compared to nonregular use, frequent use the internet for instrumental activities can significantly protect against cognitive decline. Study and work provide cognitive stimuli, and the internet serves as a platform for a cognitive reserve that continually promotes recovery from functional brain damage to prevent or delay the development of cognitive impairment [17,30]. Using the internet to study and work exercises the brain's ability to collect and process information, increasing the brain's cognitive reserve, which is associated with better cognitive performance; in addition, the information available on the internet enhances the health literacy of middle-aged and older adults. This study's results are similar to those obtained in other settings [31]. The WHO has shown the positive impact of good health literacy on mitigating the risk of cognitive decline and dementia in its Guidelines for Mitigating Cognitive Decline and Dementia.

In the study by Kobayashi et al [31], social participation was found to have three components: civic, entertainment, and cultural. It was shown that fair social participation improves the health literacy of middle-aged and older adults. After refining its purpose, we believe that using the internet for socializing and entertainment is a form of social engagement. Socializing and entertainment help people meet their emotional needs. It has been shown that from the perspective of brain evolution and functional development, the need for emotion contributes to the generation of neural networks for specific phenotypes of individual social cognition [32]. As there are many ways to use the internet for entertainment, older people can choose to enrich their lives in interesting and appropriate ways, and an enjoyable experience could bring them satisfaction and a sense of accomplishment. Regarding socializing, previous studies have shown that increased contact with family and friends via the internet by older adults has a positive impact on enhancing both life satisfaction and mood; this is especially true for older adults with health problems and limited ability to perform, since increased contact provides important opportunities to establish and maintain intimate relationships through social networking [33]. In contrast, this study showed that moderate socializing was significantly associated with reduced cognitive decline. Excessive online socializing may be related to an addiction to the emotional connections of the virtual world and a gap in real life, leading to a failure to adapt to stress. This is similar to the results of some studies, for example, Nie et al [34], who concluded that frequency of internet use is significantly negatively related to happiness, and that excessive use of the internet instead of other offline activities produces a negative effect.

Taking part in commercial-related activities was the only internet activity that showed a negative relationship with cognitive function. Internet shopping behavior was chosen for the measurement of commercial-related activities. A study by Holtfreter et al [35] showed that middle-aged and older adults are more likely to be at risk of online fraud, especially those with low self-control who are exposed to a high-risk, online, commercial-related activity environment where both online fraud and identity theft endanger the mental health of middle-aged and older adults. We hypothesized that because commercial-related activities involve money, older adults would be stressed by their infrequent and unskilled online shopping activities. Also, as the main victims of online fraud, older adults would be more cautious and worried about online payments. These emotions are detrimental to the mental health of older adults. The findings showed that when older adults used the internet frequently for commercial-related activities, the adverse effects decreased and were no longer significantly associated with negative cognitive function. We believe that older adults who do not use the internet frequently for commercial-related activities have lower digital literacy. Older adults with lower health literacy are vulnerable to online scams leading to stress adaptation failure. In contrast, older adults who frequently use the internet for commercial-related activities have higher digital literacy and good stress adaptation. Therefore, frequent use of the internet by older adults for commercial-related activities is not associated with a change in cognitive function.

In conclusion, our findings provide some evidence that the use of the internet is essential for middle-aged and older adults, especially for patients with cognitive decline. With the internet as an increasingly popular technological tool whose use is escalating, we can promote its use in the health field for the middle-aged and older adult population. Our findings advance the possibility of internet interventions for cognitive decline. Our findings also allow future research studies to fully consider the differential impact of the type, frequency, and diversity of internet use and to provide practical implementation options for nonpharmacological interventions for cognitive decline.

Strengths and Limitations

To the best of our knowledge, this study is the first to use a fixed-effects analysis to measure the relationship between internet use and cognitive decline among middle-aged and older adult Chinese people, ruling out potential endogeneity. This study confirms that greater diversity and higher frequency of internet use are associated with alleviating cognitive decline by assessing the net association between changes in internet use and changes in cognitive function. At the same time, this is one of the few studies that has finely delineated the internet use variable to explore the differential relationships of different internet use purposes on cognitive decline. These findings have practical implications and can serve as quantitative criteria to guide the design of intervention programs. The results of this study also remind subsequent researchers to discuss internet use in a disaggregated manner when exploring the causal relationship between internet use and cognitive decline.

The findings should be interpreted with caution because of the following limitations. First, some short-term benefits may diminish over time, given the impact of internet use on cognitive decline over the same period. Second, considering that there were 2 years between each wave of the CFPS study, this study used internet use and cognitive decline in both waves, rather than a lagged model. The sample size in this study is indeed more extensive than in the lagged model. However, the exact causal relationship remains to be further investigated. Third, we can only speculate about the mechanisms applicable to the findings through the observed statistical results, and the observational nature of the study limits our ability to confirm the causal and theoretical mechanisms of internet use and cognitive decline. Fourth, vocabulary is a cognitive skill; however, it is known to be relatively insensitive to cerebral pathology and, thus, unlikely to be a good choice of measures to assess cognitive decline over time. Also, the original questionnaire did not address the duration of internet use. Follow-up research can explore the effects of specific internet usage time. Although our study had a superficial discussion of the topic of mental health, the relationship between cognitive decline, internet use, and mental health was not clarified; in addition, variables measuring the mental health of middle-aged and older adults were not used in the study, and the relationship between the three needs to be further investigated.

Conclusions

Previous studies using CFPS data have demonstrated that internet use is associated with better mental health in older adults. This study refined the relationship between type, diversity, and frequency of internet use and cognitive decline by expanding internet use dimensions. In summary, there were three main findings from this study. First, using the internet for study, work, entertainment, socializing, and commercial-related activities was associated with varying degrees of alleviated cognitive decline. Second, more frequent use of the internet for study and work was associated with better outcomes, while moderate use of the internet for socializing, entertainment, and commercial-related activities was associated with better outcomes. Finally, there is a relationship between frequency and diversity of internet use and better cognitive function, with both depth and breadth of internet use being important, and with more varied and frequent use of the internet being associated with better outcomes. These findings suggest that different levels of internet use have different relationships with cognitive function in middle-aged and older adults; the findings serve as a reminder of the need for more targeted and quantitatively sound policies when designing nonpharmacological interventions for cognitive decline. Future research directions could target studies of the mechanisms responsible for the differential impact outcomes of internet use.

Acknowledgments

The authors thank the participants for their time and contributions to this research. This work was supported by the Fundamental Research Funds for the Central Universities (No. 2020WKYXQN019).

Conflicts of Interest

None declared.

References

1. Singh-Manoux A, Kivimaki M, Glymour MM, Elbaz A, Berr C, Ebmeier KP, et al. Timing of onset of cognitive decline: Results from Whitehall II prospective cohort study. BMJ 2012 Jan 05;344:d7622 [FREE Full text] [doi: 10.1136/bmj.d7622] [Medline: 22223828]
2. Prince MJ, Wu F, Guo Y, Gutierrez Robledo LM, O'Donnell M, Sullivan R, et al. The burden of disease in older people and implications for health policy and practice. Lancet 2015 Mar 07;385(9967):549-562. [doi: 10.1016/S0140-6736(14)61347-7] [Medline: 25468153]

3. Ding D, Zhao Q, Guo Q, Meng H, Wang B, Luo J, et al. Prevalence of mild cognitive impairment in an urban community in China: A cross-sectional analysis of the Shanghai Aging Study. Alzheimers Dement 2015 Mar;11(3):300-309.e2. [doi: 10.1016/j.jalz.2013.11.002] [Medline: 24613707]
4. Goh JOS. Functional dedifferentiation and altered connectivity in older adults: Neural accounts of cognitive aging. Aging Dis 2011 Jan 01;2(1):30-48 [FREE Full text] [Medline: 21461180]
5. Goh JO, Park DC. Neuroplasticity and cognitive aging: The scaffolding theory of aging and cognition. Restor Neurol Neurosci 2009;27(5):403. [doi: 10.3233/rnn-2009-0493]
6. Brookmeyer R, Johnson E, Ziegler-Graham K, Arrighi HM. Forecasting the global burden of Alzheimer's disease. Alzheimers Dement 2007 Jul;3(3):186-191. [doi: 10.1016/j.jalz.2007.04.381] [Medline: 19595937]
7. Cicerone KD, Goldin Y, Ganci K, Rosenbaum A, Wethe JV, Langenbahn DM, et al. Evidence-based cognitive rehabilitation: Systematic review of the literature from 2009 through 2014. Arch Phys Med Rehabil 2019 Aug;100(8):1515-1533. [doi: 10.1016/j.apmr.2019.02.011] [Medline: 30926291]
8. Langa KM, Levine DA. The diagnosis and management of mild cognitive impairment: A clinical review. JAMA 2014 Dec 17;312(23):2551-2561 [FREE Full text] [doi: 10.1001/jama.2014.13806] [Medline: 25514304]
9. Anttila T, Helkala E, Viitanen M, Kåreholt I, Fratiglioni L, Winblad B, et al. Alcohol drinking in middle age and subsequent risk of mild cognitive impairment and dementia in old age: A prospective population based study. BMJ 2004 Oct 04;329(7465):539 [FREE Full text] [doi: 10.1136/bmj.38181.418958.BE] [Medline: 15304383]
10. Cervilla JA, Prince M, Mann A. Smoking, drinking, and incident cognitive impairment: A cohort community based study included in the Gospel Oak project. J Neurol Neurosurg Psychiatry 2000 May;68(5):622-626 [FREE Full text] [doi: 10.1136/jnnp.68.5.622] [Medline: 10766894]
11. Verhaegen P, Borchelt M, Smith J. Relation between cardiovascular and metabolic disease and cognition in very old age: Cross-sectional and longitudinal findings from the Berlin Aging Study. Health Psychol 2003 Dec;22(6):559-569. [doi: 10.1037/0278-6133.22.6.559] [Medline: 14640852]
12. Leibovici D, Ritchie K, Ledésert B, Touchon J. The effects of wine and tobacco consumption on cognitive performance in the elderly: A longitudinal study of relative risk. Int J Epidemiol 1999 Mar;28(1):77-81. [doi: 10.1093/ije/28.1.77] [Medline: 10195668]
13. Logsdon RG, Pike KC, McCurry SM, Hunter P, Maher J, Snyder L, et al. Early-stage memory loss support groups: Outcomes from a randomized controlled clinical trial. J Gerontol B Psychol Sci Soc Sci 2010 Dec;65(6):691-697 [FREE Full text] [doi: 10.1093/geronb/gbq054] [Medline: 20693265]
14. Maselko J, Sebranek M, Mun MH, Perera B, Ahs J, Ostbye T. Contribution of generative leisure activities to cognitive function in elderly Sri Lankan adults. J Am Geriatr Soc 2014 Oct;62(9):1707-1713 [FREE Full text] [doi: 10.1111/jgs.12985] [Medline: 25139145]
15. Fernández-Blázquez MA, Ávila-Villanueva M, Maestú F, Medina M. Specific features of subjective cognitive decline predict faster conversion to mild cognitive impairment. J Alzheimers Dis 2016 Mar 08;52(1):271-281. [doi: 10.3233/JAD-150956] [Medline: 27060949]
16. Dodge HH, Zhu J, Mattek NC, Bowman M, Ybarra O, Wild KV, et al. Web-enabled conversational interactions as a means to improve cognitive functions: Results of a 6-week randomized controlled trial. Alzheimers Dement (N Y) 2015 May;1(1):1-12 [FREE Full text] [doi: 10.1016/j.trci.2015.01.001] [Medline: 26203461]
17. Jin Y, Jing M, Ma X. Effects of digital device ownership on cognitive decline in a middle-aged and elderly population: Longitudinal observational study. J Med Internet Res 2019 Jul 29;21(7):e14210 [FREE Full text] [doi: 10.2196/14210] [Medline: 31359864]
18. Department of Ageing and Life Course, World Health Organization. China Country Assessment Report on Ageing and Health. Geneva, Switzerland: World Health Organization; 2015. URL: https://www.who.int/publications/i/item/9789241509312 [accessed 2021-12-31]
19. Sokolov A, Collignon A, Bieler-Aeschlimann M. Serious video games and virtual reality for prevention and neurorehabilitation of cognitive decline because of aging and neurodegeneration. Curr Opin Neurol 2020 Apr;33(2):239-248. [doi: 10.1097/WCO.0000000000000791] [Medline: 32073439]
20. Xavier AJ, d'Orsi E, de Oliveira CM, Orrell M, Demakakos P, Biddulph JP, et al. English Longitudinal Study of Aging: Can internet/e-mail use reduce cognitive decline? J Gerontol A Biol Sci Med Sci 2014 Oct;69(9):1117-1121 [FREE Full text] [doi: 10.1093/gerona/glu105] [Medline: 25116923]
21. Myhre JW, Mehl MR, Glisky EL. Cognitive benefits of online social networking for healthy older adults. J Gerontol B Psychol Sci Soc Sci 2017 Oct 01;72(5):752-760. [doi: 10.1093/geronb/gbw025] [Medline: 26984523]
22. Yu RP, McCammon RJ, Ellison NB, Langa KM. The relationships that matter: Social network site use and social wellbeing among older adults in the United States of America. Ageing Soc 2015 Jul 02;36(9):1826-1852. [doi: 10.1017/s0144686x15000677]
23. China Family Panel Studies. URL: http://www.isss.pku.edu.cn/cfps/ [accessed 2021-12-14]
24. Huang G, Xie Y, Xu H. Cognitive ability: Social correlates and consequences in contemporary China. Chin Sociol Rev 2015;47(4):287-313 [FREE Full text] [doi: 10.1080/21620555.2015.1032161] [Medline: 27570709]

25. Valenzuela MJ, Sachdev P, Wen W, Chen X, Brodaty H. Lifespan mental activity predicts diminished rate of hippocampal atrophy. PLoS One 2008 Jul 09;3(7):e2598 [FREE Full text] [doi: 10.1371/journal.pone.0002598] [Medline: 18612379]
26. Stern Y. Cognitive reserve. Neuropsychologia 2009 Aug;47(10):2015-2028 [FREE Full text] [doi: 10.1016/j.neuropsychologia.2009.03.004] [Medline: 19467352]
27. Stern Y. What is cognitive reserve? Theory and research application of the reserve concept. J Int Neuropsychol Soc 2002 May 16;8(3):448-460. [doi: 10.1017/s1355617702813248]
28. Fratiglioni L, Paillard-Borg S, Winblad B. An active and socially integrated lifestyle in late life might protect against dementia. Lancet Neurol 2004 Jul;3(6):343-353. [doi: 10.1016/S1474-4422(04)00767-7] [Medline: 15157849]
29. Prado CE, Crowe SF. Corticosteroids and cognition: A meta-analysis. Neuropsychol Rev 2019 Sep;29(3):288-312. [doi: 10.1007/s11065-019-09405-8] [Medline: 31129865]
30. Zuelsdorff ML, Engelman CD, Friedman EM, Koscik RL, Jonaitis EM, Rue AL, et al. Stressful events, social support, and cognitive function in middle-aged adults with a family history of Alzheimer's disease. J Aging Health 2013 Oct;25(6):944-959 [FREE Full text] [doi: 10.1177/0898264313498416] [Medline: 23945762]
31. Kobayashi LC, Wardle J, von Wagner C. Internet use, social engagement and health literacy decline during ageing in a longitudinal cohort of older English adults. J Epidemiol Community Health 2015 Mar;69(3):278-283 [FREE Full text] [doi: 10.1136/jech-2014-204733] [Medline: 25428933]
32. Lesch K. Linking emotion to the social brain. The role of the serotonin transporter in human social behaviour. EMBO Rep 2007 Jul;8 Spec No:S24-S29 [FREE Full text] [doi: 10.1038/sj.embor.7401008] [Medline: 17726438]
33. Heo J, Chun S, Lee S, Lee KH, Kim J. Internet use and well-being in older adults. Cyberpsychol Behav Soc Netw 2015 May;18(5):268-272. [doi: 10.1089/cyber.2014.0549] [Medline: 25919967]
34. Nie P, Sousa-Poza A, Nimrod G. Internet use and subjective well-being in China. Soc Indic Res 2016 Jan 4;132(1):489-516. [doi: 10.1007/s11205-015-1227-8]
35. Holtfreter K, Reisig MD, Pratt TC, Holtfreter RE. Risky remote purchasing and identity theft victimization among older internet users. Psychol Crime Law 2015 Apr 08;21(7):681-698. [doi: 10.1080/1068316x.2015.1028545]

Abbreviations

CFPS: China Family Panel Studies
ELSA: English Longitudinal Study of Aging
WHO: World Health Organization

Edited by G Eysenbach; submitted 16.11.20; peer-reviewed by J Xiao, H LaMonica, W Zhang; comments to author 15.03.21; revised version received 18.05.21; accepted 18.11.21; published 24.01.22.

Please cite as:
Yu X, Mu A, Wu X, Zhou L
Impact of Internet Use on Cognitive Decline in Middle-Aged and Older Adults in China: Longitudinal Observational Study
J Med Internet Res 2022;24(1):e25760
URL: https://www.jmir.org/2022/1/e25760
doi:10.2196/25760
PMID:35072642

Original Paper

A New Remote Guided Method for Supervised Web-Based Cognitive Testing to Ensure High-Quality Data: Development and Usability Study

Victoria Leong[1,2,3], PhD; Kausar Raheel[1], MSc; Jia Yi Sim[1], BA; Kriti Kacker[1], BSc; Vasilis M Karlaftis[4], PhD; Chrysoula Vassiliu[5], BA; Kastoori Kalaivanan[2], PhD; S H Annabel Chen[1,2,3,6], PhD; Trevor W Robbins[4,7], PhD; Barbara J Sahakian[7,8], PhD; Zoe Kourtzi[4], PhD

[1]Psychology, School of Social Sciences, Nanyang Technological University, Singapore, Singapore

[2]Centre for Research and Development in Learning, Nanyang Technological University, Singapore, Singapore

[3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore, Singapore

[4]Department of Psychology, University of Cambridge, Cambridge, United Kingdom

[5]Faculty of Modern and Medieval Languages and Linguistics, University of Cambridge, Cambridge, United Kingdom

[6]National Institute of Education, Nanyang Technological University, Singapore, Singapore

[7]Behavioural and Clinical Neuroscience Institute, University of Cambridge, Cambridge, United Kingdom

[8]Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom

Corresponding Author:
Victoria Leong, PhD
Psychology
School of Social Sciences
Nanyang Technological University
48 Nanyang Avenue
Singapore, 639818
Singapore
Phone: 65 6514 1052
Email: victorialeong@ntu.edu.sg

Abstract

Background: The global COVID-19 pandemic has triggered a fundamental reexamination of how human psychological research can be conducted safely and robustly in a new era of digital working and physical distancing. Online web-based testing has risen to the forefront as a promising solution for the rapid mass collection of cognitive data without requiring human contact. However, a long-standing debate exists over the data quality and validity of web-based studies. This study examines the opportunities and challenges afforded by the societal shift toward web-based testing and highlights an urgent need to establish a standard data quality assurance framework for online studies.

Objective: This study aims to develop and validate a new supervised online testing methodology, remote guided testing (RGT).

Methods: A total of 85 healthy young adults were tested on 10 cognitive tasks assessing executive functioning (flexibility, memory, and inhibition) and learning. Tasks were administered either face-to-face in the laboratory (n=41) or online using remote guided testing (n=44) and delivered using identical web-based platforms (Cambridge Neuropsychological Test Automated Battery, Inquisit, and i-ABC). Data quality was assessed using detailed trial-level measures (missed trials, outlying and excluded responses, and response times) and overall task performance measures.

Results: The results indicated that, across all data quality and performance measures, RGT data was statistically-equivalent to in-person data collected in the lab (P>.40 for all comparisons). Moreover, RGT participants out-performed the lab group on measured verbal intelligence (P<.001), which could reflect test environment differences, including possible effects of mask-wearing on communication.

Conclusions: These data suggest that the RGT methodology could help ameliorate concerns regarding online data quality—particularly for studies involving high-risk or rare cohorts—and offer an alternative for collecting high-quality human cognitive data without requiring in-person physical attendance.

(J Med Internet Res 2022;24(1):e28368) doi:10.2196/28368

KEYWORDS

web-based testing; neurocognitive assessment; COVID-19; executive functions; learning

Introduction

Background

In 2020, the global COVID-19 pandemic brought human lab-based psychological research to an abrupt halt as social distancing measures preventing disease transmission forced the mass closure of laboratory facilities and prevented all but essential human contact, disrupting academic research at a profound level [1,2]. During this period of suspension, the research community has inexorably moved toward remote protocols to replace face-to-face activities. There has been an exponential rise in the use of online platforms such as video conferencing (Zoom [3] and Skype) and online learning [3,4] for day-to-day academic activities, and the use of social media platforms has surged, not just as a means for interacting and connecting with others, but also for participant recruitment and outreach [5]. Concomitantly, interest in online experimental alternatives to in-person cognitive testing has grown significantly [1], and there is an increasing focus on methodological developments that will allow the field to adapt to a changed world where reduced social contact is the new norm [1,6,7]. However, amidst this push toward new online research technologies, core issues of data quality and validity should not be overlooked, and simple assumptions of equivalence between lab-based and online tests (eg, on the basis that the use of similar tasks and platforms is sufficient [8-10]) should not be made. Failure to address these issues could lead to a proliferation of poorly regulated online research studies, worsening the current reproducibility crisis and raising new ethical dilemmas [11]. This study examines the opportunities and challenges afforded by the shift toward web-based testing, highlighting an urgent need to establish a standard data quality assurance framework for online studies and proposing a new supervised online testing methodology, remote guided testing (RGT), which could mitigate some of these challenges and offer an alternative for collecting high-quality human cognitive data within social distancing constraints.

The Rise of Web-Based Cognitive Testing: Opportunities and Challenges

Cognitive tests are valuable psychological tools used extensively to examine mental executive processes such as learning, decision-making, inhibition, and working memory [12-31]. Tests of executive processes have typically involved pen-and-paper administration in lab-based settings, allowing the experimenter to confirm the participant's identity, deter dishonesty, and promptly assist with queries or technical problems. The standardized testing environment and equipment further aid to ensure replicability and reproducibility of lab-based protocols [32-37]. However, this traditional in-person approach is time-consuming and highly susceptible to human error [12,32]. Further, since lab-based testing requires participants' physical attendance in the laboratory, sampling may not be population-representative [38-40]. For instance, Henrich et al [38], Nielsen et al [39], and Arnett [40] report that participants in lab-based studies consist predominantly of Western, educated, industrialized, rich, and democratic populations. An analysis of 6 major American Psychological Association journals [40,41] showed that a significant number of studies reported in these journals relied predominantly on American students. Much of the normative data on psychological and cognitive processes has been obtained from a North American, White, high socioeconomic status, and well-educated demographic, raising the possibility that neuropsychology may be insensitive to cultural and ethnic differences [42,43]. Structural racism, that is, the establishment of a series of dynamics that promote White people as the norm (to the exclusion or minimization of Black and ethnic minority people), may also have led to the routine acceptance of nonrepresentative standardization samples that are primarily White, creating false normative expectations for Black and ethnic minorities [44,45]. These biased practices in psychological assessment have long gone unchallenged, partially due to a prevalent belief in universalism (ie, the theory that cognitive processes are essentially the same across humankind, irrespective of cultural background) [46,47]. This highlights the need for a wider representation of ethnic minority groups and cultures in psychological studies. Web-based testing could help to ameliorate this gap and reach a more diverse global audience for neuropsychological research.

In the current digital age, the mass availability of personal computers and web capability affords new avenues for cognitive testing using more cost-effective, automated, and open approaches [12,48,49]. Accordingly, there is growing momentum in the use of online platforms to assess cognitive function [12,50-52]. Computerized online testing platforms such as Gorilla [50], Inquisit [52], and the Cambridge Neuropsychological Test Automated Battery (CANTAB) [51] can offer several advantages, including (1) simple and precise control of experimental parameters, (2) automatic calculation of key performance indices, (3) access to normative databases, (4) accessibility to a wider population of users (with use of crowdsourcing tools), (5) centralized and secure data storage on professional servers, and (6) relatively low administrative cost per head [32,34,48,53]. Notably, web-based online testing removes the physical constraint of test locations, permitting a much wider (and more representative) demographic reach [34,54]. Further, social media platforms, recruitment portals (eg, Amazon's MTurk and Prolific Academic), and online advertisements have broadened horizons by increasing ease of participant recruitment and enabling high throughput data collection from large populations, which is less feasible in traditional lab-based settings [49,55-58].

The COVID-19 pandemic has fueled the growth of "telehealth" or remote access to health care services, bringing to the fore particular challenges in providing remote neuropsychological assessments, psychoeducation, and rapport building [59].

Ongoing demand for telehealth services even in the postpandemic era is likely, as these may be valuable solutions when physical presence in the clinic is impossible for other reasons (eg, sickness, workload, etc). Relatedly, remote testing (or tele-testing) is becoming increasingly popular amongst clinicians. Tele-testing, often combined with face-to-face testing, results in a hybrid approach that can cater to the specific needs of patients and their families [59]. Singh and Germane [60] elaborate on this hybrid approach, which they call "hybrid neuropsychology" (HN). HN enables clinicians to effectively and steadily modernize their practice considering individuals' needs and (technological) capabilities and evaluating which tasks are ideal for online or remote administration. Indeed, HN incorporates tele-testing practices and screen-sharing options wherever materials have been digitized properly—bearing similarities to the remote guided methodology proposed in this study. Additionally, rigorous and standardized protocols for web-based test delivery are not yet available for most neuropsychological tasks, limiting the confidence with which clinicians and scientists can adopt these methodologies in daily practice. This study addresses a growing demand for remote methods of neuropsychological measurement in both research and clinical settings by providing one such detailed remote administration protocol for a suite of executive function and cognitive tasks that are highly relevant to clinical assessment.

Further, a long-standing debate still exists regarding the data quality, comparability, replicability, and validity of web-based versus traditional lab-based data collection methods [61,62]. On the one hand, direct comparisons of web-based and lab-based data samples from web pioneers such as Germine and colleagues [63-65] on a series of large-scale web-based studies on memory and perception (testmybrain.org) indicate that the reliability, replicability, and theoretical consistency of self-selected web samples are comparable to lab-collected data in terms of mean performance and performance variability, even with anonymous, uncompensated, and unsupervised participants. On the other hand, concerns have been raised, and not yet fully allayed, about the experimental rigor of web-based testing [36,37], particularly regarding the lack of control over and higher variability of hardware specifications and the test environment [36,37,66]. For instance, a study by Bauer et al [36] reported that most online studies suffer from a lack of environmental control and participant distraction. Further, most online studies do not monitor or report measures of participants' environment, their equipment specification, or web capability. These data, when reported, typically reveal large variations in the equipment and computer specifications used by participants [34,66]. In a landmark study on computing specifications in online and lab-based studies, Bridges et al [66] compared the pairings of several web-based experimental platforms, such as Gorilla and jsPsych, with different operating systems, such as Windows, macOS, and Ubuntu. In data collected from over 110,000 trials, macOS yielded the worst performance across all experimental web platforms, particularly for visual stimuli. This variability also suggests that online studies may not achieve a similar level of precision as lab-based studies. Moreover, since the data are contributed anonymously and without supervision, online data could be compromised by dishonest participants with low or questionable motivation [36].

In fact, several comparative studies report only a moderate correlation between web-based cognitive performance and its paper and pencil alternatives across different cognitive tasks and populations [32,36,67]. For instance, Backx et al [34] compared cognitive data obtained from the CANTAB platform, which was collected using unsupervised web-based and lab-based administration. Intraclass and bivariate correlations showed that several key performance indices (errors, correct trials, and response sensitivity) were highly comparable across the two settings, with intraclass correlation coefficients ranging from ρ=0.23-0.67. However, participant reaction times (RTs) were off-task significantly and consistently slower for web-based assessments, and none of the 5 RT measures that were assessed met the full criterion for comparability across settings, namely, reliability, equivalence, and agreement. Further, in the online setting, over 90% of participants reported being distracted across 5 different cognitive tasks, as compared to none in lab settings, and 2 online participant data sets were excluded due to a high number of errors on the spatial working memory (SWM) task. These statistics suggest that a poor test environment and miscomprehension of instructions could affect participant performance in online settings. Indeed, previous research suggests that a lack of incentive can make participants careless in their responses or even deceptive, as participants' identity or behavior cannot be actively monitored [68]. Further, online participants may show lower task engagement by investing less time and focus on reading task instructions, leading to higher dropout rates than in laboratory settings [69,70]. Therefore, with current unsupervised web-based testing protocols, there appears to be a trade-off between data quantity (diversity and ease of collection) and data quality [37,66,71]. Given these known pitfalls in online testing, along with increased efforts toward standardizing best practices in online task administration, data quality indices, and reporting benchmarks (eg, Feenstra et al [72]), this study addresses a timely need to develop better test protocols and data quality assurance frameworks for web-based cognitive testing, specifically addressing issues with online participant engagement.

Data Quality Assurance Framework for Web-Based Cognitive Testing

To assess data quality, it is first important to establish the indices and benchmarks by which data quality will be measured. On these points, there is currently no clear consensus. General statistics on participant noncompletion, data attrition, and technical issues show that it is common to exclude data from participants who encounter technical difficulties, display dishonesty, or fail to complete the assigned tasks [71,73-75]. However, a more sensitive test of data quality pertains to the "usable fraction" of data that remains after task-specific exclusion criteria for data cleaning have been applied. One common index used for data exclusion is RT since responses that are too fast (or too slow) are likely to reflect participant inattentiveness or task disengagement. Depending on the stimuli presented and the complexity of task demands, participant response latencies in cognitive tasks mostly vary between 400 milliseconds to 2000 milliseconds [76]. However, for web-based tasks, there is an additional (technical) source of variability to the measured response times. Collecting response latencies from

many (eg, hundreds) individual trials requires a software program to be installed on the participant's computer (ie, client-side), to present the stimuli and collect response latencies locally. This reduces the temporal variation introduced by communication across networks and server response times if each response must be sent over the network connection back to the server to be recorded. Several client-side technologies have been used to create such programs, with perhaps the most popular being JavaScript, Java, and Flash [76].

However, most of these programs introduce a small but variable delay in the recorded response times. In addition to the program itself, this delay is influenced by the computer's operating system, browser, hardware quality, and any background programs that may be running. For example, when Schubert et al [76] measured standard automated response times (ie, robot detection of a simple visual stimulus) natively using DMDX software and a keyboard, the mean response time was 68.24 milliseconds (SD 3.18). These mean latencies were higher when other programs were used, for instance, E-prime (84.58 milliseconds, SD 6.25) and Superlab (98.18 milliseconds, SD 4.17). Interestingly, mean response latencies were highly comparable for the web version of Inquisit (66.21 milliseconds, SD 2.74). When comparing human response times on a Stroop Task, Schubert et al [76] reported that DMDX-recorded responses were significantly faster (mean 551.98 milliseconds, SD 201.38) than Flash-based web software ScriptingRT (mean 631.63 milliseconds, SD 243.42), although the measured Stroop effect (difference in response latency between incongruent and congruent trials) was similar. Therefore, web-collected response latency data should be carefully handled as the measured timings may be impacted by both psychological (eg, participant distraction and inattentiveness) and technical factors, although the latter issue is ameliorated to some extent by newer and better online experimental platforms.

Some web-based studies implement a hard cut-off to exclude response latencies past a particular threshold to optimize data quality. For instance, Kim et al [77] excluded outlier responses that were faster than 300 milliseconds or slower than 5000 milliseconds in a psycholinguistic task they employed. However, these excluded trials represented a mere 0.75% of their data (for both lab-based and web-based cohorts) which could either suggest superb data quality or that their latency criteria were too lax for the particular task. Similarly, Eisenberg et al [55] excluded participants whose median response latencies were shorter than 200 milliseconds across a wide range of cognitive tasks. They also implemented three additional quality checks: (1) <25% omitted (missed) responses, (2) >60% task accuracy, and (3) no single response given >95% of the time. However, although rates for participant noncompletion and multiple task failure were reported (Table 1), the number of trials and data sets that failed their other response latency, omission, and distribution quality checks were not reported. Further, for some tasks (Stop Signal, probabilistic selection, and two-step decision tasks), the data sets had between 10% to 30% missing values that were identified through additional quality control (manipulation) checks. Adding this figure to the reported 21% of data exclusions suggests that the actual fraction of "usable data" could be as low as 50% on some web-based tasks.

Table 1 provides examples of other data exclusion statistics for analogous lab-based and web-based cognitive studies. These statistics report participant completion/dropout rates rather than trial-level data quality indices. Studies by Hicks et al [79] and Ruiz et al [80], who administered matched sets of working memory and declarative memory tasks in lab-based and online conditions, show a clear and consistent trend toward a higher rate of data exclusion and noncompletion for web-based testing, with typically 15% to 20% more online participants excluded for noncompletion (dropout) and technical issues. Participant dropout issues appear to be particularly exacerbated for online cognitive training studies, with one study reporting that 32% (80/249) of initially recruited healthy older adult participants eventually withdrew from their 12-week cognitive flexibility web-based training study [81]. However, this brief scan of the literature also highlights that, except for a few studies (eg, Backx et al [34]), little is reported about what occurs during online experimental testing or about the computing and web capabilities of participants, and few benchmarks exist for identifying and removing poor quality web-based data, beyond crude RT thresholds and major task failures. There is a clear need to develop standardized data quality indices that web-based studies should collect and report, including the recommended benchmark(s) for these indices.

Table 1. Examples of data exclusion statistics reported for lab-based and web-based cognitive studies.

Study type and citation	Task(s)	Data excluded
Lab-based		
Kim et al [77]	Lab, psycholinguistic task	5/42 (11.9%) participants excluded for high error rates or being outside demographic. Reaction time outlier removal=0.75% of total data
Von Gunten et al [82]	Lab, inhibition tasks (antisaccade, go/no go, and Stop Signal)	37/463 (7.99%) participants excluded
Backx et al [34]	Lab, CANTAB[a] tasks[b]	No exclusions, no distractions reported
Hicks et al [79]	Lab (experiments 1 and 3), working memory tasks	Experiment 1: 0/58 (0%) participants excluded, although 10% of participants reported cheating; experiment 3: 10/112 (8.9%) participants excluded due to excessive missing data
Ruiz et al [80]	Lab, working memory[c], nondeclarative/declarative memory tasks	(a) OSpan[d], 0% excluded; (b) MLAT5[e], 0% excluded; (c) CVMT[f], 1/50 (2%) participants excluded
Baniqued et al [78]	Cognitive video training	27/219 (12.3%) participants excluded or withdrew
Web-based		
Kim et al [77]	Online, psycholinguistic task	3/39 (7.7%) participants excluded for high error rates or being outside demographic. Reaction time outlier removal=0.75% of total data
Eisenberg et al [55]	Online (using Amazon Turk), inhibition tasks (go/no go, Stop Signal)	102/662 (15.4%) participants excluded for noncompletion of task battery; 38/560 (6.8%) participants further excluded for failing 4 or more tasks
Backx et al [34]	Online, CANTAB tasks	2/18 (11.1%) participants excluded, high SWM[g] errors; distractions: 16/18 participants for PAL[h], ERT[i], OTS[j], and PRM-I[k]; 17 participants for SWM, RVP[l], PRM-D[m]
Hicks et al [79]	Online (experiments 2 and 4), working memory tasks	Experiment 2: 12/100 (12%) participants excluded for failure to complete test battery within 24 hours; Experiment 4: 28/112 (25%) participants excluded due to noncompletion of task battery
Ruiz et al [80]	Online, working memory, nondeclarative/declarative memory tasks	(a) OSpan, 7/50 (14%) participants excluded; (b) MLAT5, 8/15 (16%) participants excluded; (c) CVMT, 10/50 (20%) participants excluded
Buitenweg et al [81]	Cognitive flexibility training	91/249 (36.5%) participants excluded for not meeting criteria (N=11) or withdrew from study (N=80)

[a]CANTAB: Cambridge Neuropsychological Test Automated Battery

[b]CANTAB tasks include SWM, PAL, ERT, OTS, PRM-I, RVP, and PRM-D.

[c]Memory tasks include OSpan, MLAT, and CVMT.

[d]OSpan: automated operation span task.

[e]MLAT: modern language aptitude test.

[f]CVMT: continuous visual memory task.

[g]SWM: spatial working memory.

[h]PAL: paired associates learning.

[i]ERT: emotion recognition task.

[j]OTS: one touch stockings of Cambridge.

[k]PRM-I: pattern recognition memory-immediate.

[l]RVP: rapid visual processing.

[m]PRM-D: pattern recognition memory-delayed.

A New Supervised Online Method: Remote Guided Testing

In the preceding sections, we discussed the promise of web-based cognitive testing, specifically its scalability and reach, and the current challenges for data quality assurance. In part, questions over experimental rigor and data quality have arisen due to the unsupervised nature of web-based testing [36,37,66]. Without supervision, experimenters have no control over (or insight into) the test environment and no way to monitor participant performance, deter dishonesty, or influence participant motivational and attentional states during task performance. Further, even genuinely motivated participants may struggle with tasks that have complex instructions and misunderstand what is required of them, leading to wasted effort and unusable data. Finally, without a human experimenter on hand to troubleshoot problems, participants experiencing technical issues may become frustrated and stressed, leading to poorer motivation and performance. To bridge this gap, we propose here a new method of supervised online data collection,

RGT. This hybrid method marries the convenience and reach of online web-based testing with the enhanced rigor and quality control of in-person lab-based data collection. The addition of a supervisory component, including greater environmental control, aims to mitigate data quality degradation and attrition due to psychological or technical factors.

The RGT method simulates lab-based experimental testing via a video conferencing platform. Similar to in-person testing, the experimenter arranges to meet the participant online at a specific date and time and guides the participant virtually through each step of the experimental process. This includes obtaining informed consent, providing technical support for software installation, troubleshooting problems, monitoring performance, providing feedback where appropriate, and debriefing. The experimenter also helps the participant to optimize their test environment (including lighting, sound, and minimizing distractions) and collects detailed data about the hardware, software, and web capabilities of the participant. Additionally, the remote tester can schedule comfort breaks (for toilet trips, food or drink, rest, exercise) so as not to affect test delivery or data collection adversely. This method is novel in its holistic approach as it provides a fully supervised and interactive online test experience, which to our knowledge has not been reported before for web-based cognitive testing.

To provide a deeper evaluation of the RGT method on data quality, we measure and report 3 trial-level data quality indicators across a range of web-based cognitive tasks: (1) missed responses, (2) data exclusions (at both trial and participant levels), and (3) RTs. To ensure close comparability and to isolate the effect of test modality, participants completed identical web-based versions of each cognitive task either in-person in a psychology lab or at home via RGT. In both conditions, participants received expert supervision while they completed a range of cognitive tasks assessing executive function (cognitive flexibility, inhibition, and working memory) and learning. While most of these tasks rely on measures of accuracy, we specifically included tasks with RT-dependent outcome measures, such as the Stroop Task [20] and the Stop Signal Task [21,22]. Given that there are well-quantified effects of web-based testing in terms of slower participant RTs on cognitive tasks [34,76], we assessed if (and the extent to which) these differences could be ameliorated through greater supervisory and environmental control. Finally, to increase the generalizability of our findings, 3 different web-based experimental platforms were tested; CANTAB, Inquisit, and i-ABC. We hypothesized that the inclusion of supervision via the RGT method would yield high fidelity cognitive data that match lab-collected cognitive data in all measured indices of data quality and task performance (including RTs).

Methods

Participants

A total of 85 healthy Singaporean young adults participated in the study and contributed data face-to-face (F2F; n=41) and via RGT (n=44). A further 4 RGT and 5 F2F participants had initially expressed interest but subsequently withdrew from the study. Data from these participants were not included in any analyses. All participants were native English speakers, reported no history of clinically diagnosed mental illness or developmental difficulties, and had normal or corrected hearing and vision. Recruitment was conducted through online advertisements in social media outlets and through the University's recruitment channel. The demographic information for both groups is detailed in Table 2. A two-tailed *t* test confirmed that there was no significant difference in age between groups (t_{83}=–1.29, *P*=.20), and gender, ethnicity, education, and income distributions were also similar.

All 85 participants attended and completed their scheduled testing session(s). None of them withdrew midway through their session(s). All 44 (100%) participants in the RGT group completed all 10 computerized tasks on web-based platforms. However, only 22 (53.7%) F2F participants completed all the computerized tasks on web-based platforms (17 females and 5 males; mean age 21.06 years, range=18.11-26.68 years, SD 2.09 years). The remaining 19 (46.3%) F2F participants were tested before COVID-19 lockdown restrictions and therefore only completed the 3 i-ABC tasks, vocabulary, and Digit Span tasks in a format similar to the other participants. The other Inquisit tasks (Trails, Stop Signal, and Stroop) had either been completed on paper or using a different (offline) platform, and the CANTAB tasks were not administered. As these task-related differences could have generated performance differences, for consistency, only the data from the i-ABC tasks, vocabulary, and Digit Span were analyzed for these 19 (46.3%) F2F participants.

Table 2. Summary of participant demographics by testing modality.

Demographic variable	Modality (group)		
	F2F[a] (n=41)	RGT[b] (n=44)	Total (N=85)
Age (years)			
Mean (SD)	21.54 (2.26)	22.14 (2.05)	21.85 (2.16)
Range	18.11-29.22	18.51-26.83	18.11-29.22
Gender, n (%)			
Female	29 (70.7)	33 (75)	62 (72.9)
Male	12 (29.3)	11 (25)	23 (27.1)
Ethnicity, n (%)			
Chinese	34 (82.9)	36 (81.8)	70 (82.4)
Malay	4 (9.8)	6 (13.6)	10 (11.8)
Indian	2 (4.9)	2 (4.5)	4 (4.7)
Not reported	1 (2.4)	0 (0)	1 (1.2)
Income by dwelling, n (%)			
Lower	13 (31.7)	16 (36.4)	29 (36.3)
Higher	24 (58.5)	27 (61.4)	51 (63.7)
Not reported	4 (9.8)	1 (2.3)	5 (5.9)
Highest education level, n (%)			
Secondary School	27 (65.9)	23 (52.3)	50 (58.8)
Bachelor's Degree	12 (29.3)	16 (36.4)	28 (32.9)
Not reported	2 (4.9)	5 (11.4)	7 (8.2)
Handedness, n (%)			
Right-handed	38 (92.7)	42 (95.5)	80 (94.1)
Left-handed	2 (4.9)	2 (4.5)	4 (4.7)
Not reported	1 (2.4)	0 (0)	1 (1.2)

[a]F2F: face-to-face.

[b]RGT:remote guided testing.

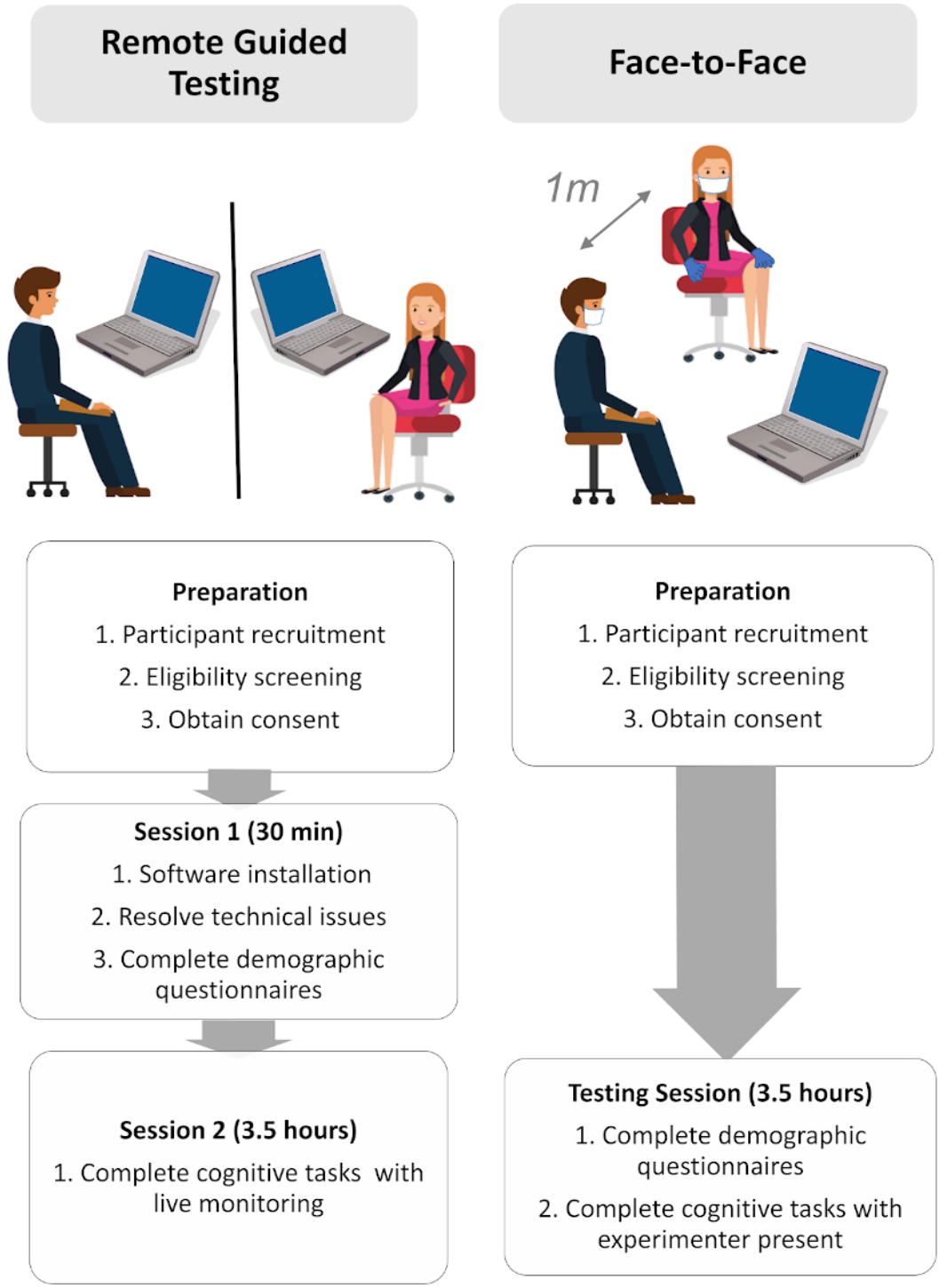
Equipment

For the F2F group, experimental testing was conducted in a psychology lab using a standard testing laptop (HP ProBook 430 G2/G3, Intel Core i7 2/2.4GHz, 8 GB RAM, 500 GB HDD +256 GB SSD, 13.3" display) running Windows 10 OS (Microsoft Corporation), with a wired mouse. For the RGT group, sessions were completed at home using participants' personal laptops or desktops that had to meet certain minimum requirements (Multimedia Appendix 1). To assess the actual quality of their computing hardware and web capability, all RGT participants completed an equipment questionnaire (Multimedia Appendix 2).

Procedure

A standard operating procedure (Multimedia Appendix 1) was followed to ensure standardized methodology and task delivery for participants in the F2F and RGT groups. Prevailing COVID-19 precautions such as mask-wearing, temperature-taking, and checking of travel/quarantine history were also applied. The study protocol was approved by the NTU Ethics Institutional Review Board (IRB-2020-02-001). In brief, F2F participants completed 1 single in-person testing session lasting 3.5 hours whereas, RGT participants completed 2 separate online video-conferencing sessions, which were conducted via secure Microsoft Teams or Zoom meetings. Both online sessions were video recorded and lasted 4 hours in total. During the first 30-minute install and set-up session, participants were guided by the experimenter to download, install, and test all necessary software. The testing environment was assessed to provide recommendations for minimizing noise and disruption (Multimedia Appendix 3, see testing environment checklist), and computing and input devices were recorded (Multimedia Appendix 2). During the second 3.5-hour session, RGT participants performed all computerized tests under the supervision of the experimenter, who remained online throughout the session (with their video off/muted where appropriate). Five different task orders were generated to ensure that no 2 tasks from the same cognitive domain were administered consecutively. The respective procedures are summarized in Figure 1.

Figure 1. Overview of remote guided and face-to-face testing processes.



Tasks

Overview

In total, participants completed 10 online experimental tasks assessing aspects of executive functioning (cognitive flexibility, working memory, and inhibitory control), learning, and verbal intelligence. These tasks were delivered using 3 different experimental web platforms: i-ABC [83], Inquisit 5 (Millisecond Software), and CANTAB (Cambridge Cognition); or delivered verbally by the experimenter, as summarized in Table 3. Both groups also completed a short online demographics questionnaire.

Table 3 provides an overview of the various delivery platforms that were used in this study and the full set of tasks. Checkmarks indicate the respective platform on which the task was administered.

Table 3. Summary of experimental tasks administered and respective delivery platforms.

Domains and tasks	Delivery platform			
	i-ABC	CANTAB[a]	Inquisit	Verbal
Cognitive flexibility				
WCST[b]	✓	—[c]	—	—
PR[d]	✓	—	—	—
TMT[e]	—	—	✓	—
IED[f]	—	✓	—	—
Working memory				
SWM[g]	—	✓	—	—
WAIS-IV BDS[h]	—	—	—	✓
Inhibition				
Stroop Task (Stroop)	—	—	✓	—
SST[i]	—	—	✓	—
Learning				
SL[j]	✓	—	—	—
Verbal IQ[k]				
WASI-II[l] vocabulary (vocab)	—	—	—	✓

[a]CANTAB: Cambridge Neuropsychological Test Automated Battery.

[b]WCST: Wisconsin Card Sort Test.

[c]Empty cells indicate that the particular task was not administered via the specific delivery platform.

[d]PR: probabilistic learning and reversal.

[e]TMT: trail making task.

[f]ED: intra-extra dimensional set shift.

[g]SWM: spatial working memory.

[h]WAIS-IV BDS: Weschler Adult Intelligence Scale–Fourth Edition Backwards Digit Span.

[i]SST: Stop Signal Task.

[j]SL: structure learning.

[k]Q: intelligence quotient.

[l]WASI-II: Weschler Abbreviated Scale of Intelligence–Second Edition.

i-ABC Platform

Three experimental tasks were administered on the i-ABC platform [83]. The Wisconsin Card Sort Test [15,16] and the probabilistic reversal task [84] were measures of cognitive flexibility, whilst the Structure Learning task [19,83] assessed statistical learning. The i-ABC website enabled the administration of the 3 tasks on a platform that simulated playing a “space-themed” video game, and participants earned points for completing the tasks. Detailed task descriptions and performance indices are provided in Multimedia Appendix 4 (see subsections 1-3).

Inquisit 5 Web Platform

Computerized versions of the Stroop Task, Stop Signal Task, and the Trail Making Task were hosted and administered on the Inquisit 5 web player by Millisecond software [85]. The software was downloaded before the session, and when each task link was opened, participants were prompted to key in their unique ID before launching into full-screen mode. The display dimensions of the task stimuli were standardized and automatically adjusted by the software according to the computer physical screen display size. Detailed task descriptions and performance indices are provided in Multimedia Appendix 4 (see subsections 4-6).

CANTAB Platform

The intra-extra dimensional (IED) set shift task and spatial working memory task were both administered as part of CANTAB [26,27,86]. Detailed task descriptions and performance indices are provided in Multimedia Appendix 4 (see subsections 7-8).

Verbal Delivery

The vocabulary subtest of the second edition of the Wechsler Abbreviated Scale of Intelligence (WASI-II) [28] and the Backwards Digit Span subtest from the fourth edition of the Weschler Adult Intelligence Scale (WAIS-IV) [29] were

administered via verbal delivery to assess verbal intelligence and verbal working memory respectively. Detailed task descriptions and performance indices are provided in Multimedia Appendix 4 (see subsections 9-10).

Data Quality Indicators

Missed Trials

For tasks involving a response within a specified time limit, the number of missed trials was calculated. If a participant did not enter a response within the specified time limit for a trial, this was considered a missed trial. As some tasks (eg, Stroop) required a response before proceeding, this index was not available for these tasks.

Data Exclusion

Participant data could be excluded either at the trial level or at the task level (ie, all participant data removed for that task).

Trial-Level Exclusions (Outliers)

Single trials were excluded if the RT on that trial was outlying (either too fast or too slow). Referencing previous research using similar tasks [55,77], response times faster than 300 milliseconds are generally deemed to indicate participant inattentiveness or a failure to fully process the stimulus on that trial. Conversely, response times that are greater than SD 2.5 of the response time distribution are also generally considered to be outliers, indicating failures of attention. The F2F RT distribution was used to set a fixed threshold for both groups to ensure that the basis for identifying slow outlier response times was consistent and provided a fair basis of comparison of data quality. Specifically, slow outliers were defined as RTs >2.5 SDs above the mean of the F2F distribution for each task. Any trials with RTs slower than this threshold (for both F2F and RGT participants) were considered outliers and removed.

As an exception for the Stop Signal Task, and with reference to Verbruggen et al [22], the first trials for each of the 3 blocks were removed as participants were not expected to have fully engaged with the task at this early stage. Additionally, RTs under 300 milliseconds were not removed for the Stop Signal Task as participants were required to provide a speeded response, and failures of inhibition were of core interest. As the CANTAB web platform did not provide trial-level data, no trials were excluded for the IED or SWM tasks. Since the verbal delivery tasks (WASI Vocabulary and Backwards Digit Span) were administered manually by the experimenter in both the F2F and RGT settings, they were not subject to trial-level exclusions.

Task-Level Exclusions

At the task level, participant data were excluded for either technical or performance reasons. Data were excluded for technical reasons if the participant experienced difficulties with the experimental platform, equipment, or testing environment during that task. Task-level performance exclusions occurred if the participant's total number of missed and excluded trials was >25% of all trials for that task (see previously discussed criteria for trial-level exclusions).

Reaction Times

The final index of data quality was the mean RT (by participant) of the remaining included trials. This was used as a data quality indicator because previous studies have indicated that mean RTs may be more variable/longer during web-based delivery of experimental tasks [34,76]. As the CANTAB web platform did not provide trial-level data, this RT index was not available for the IED or SWM tasks.

Results

Technology Profile of Remote Guided Participants

Each RGT participant completed the experimental tasks at home using their personal computer and internet connection. Although all participants used equipment that met certain minimum standards as stated in the eligibility criteria (see Equipment in the Methods section), we wished to determine the actual range and quality of technology that was being used. As shown in Figures 2-3 and detailed in Table 4, the lab equipment was a close match to the hardware specifications reported by the RGT group (eg, Windows OS, Intel Core i7 processor, 13-inch screen, 1920 x 1080 resolution, approximately 8 GB RAM). In terms of web capability, most RGT participants had better internet download/upload speeds than the F2F group (mean of 77.9/70.4 Mb/s vs 44.6/48.1 Mb/s) but slightly longer internet latencies (mean of 10.6 milliseconds as compared to 5 milliseconds).

Figure 2. Hardware specifications for remote guided participants (total N=44), including computer (a) brand; (b) operating system; (c) screen size (in inches) (d) screen resolution (in pixels); (e) processor and (f) RAM (in GB).

Figure 3. Web capability for remote guided participants (total n=44), including (a) internet download/upload speed (higher=better); and (b) internet latency (shorter=better).

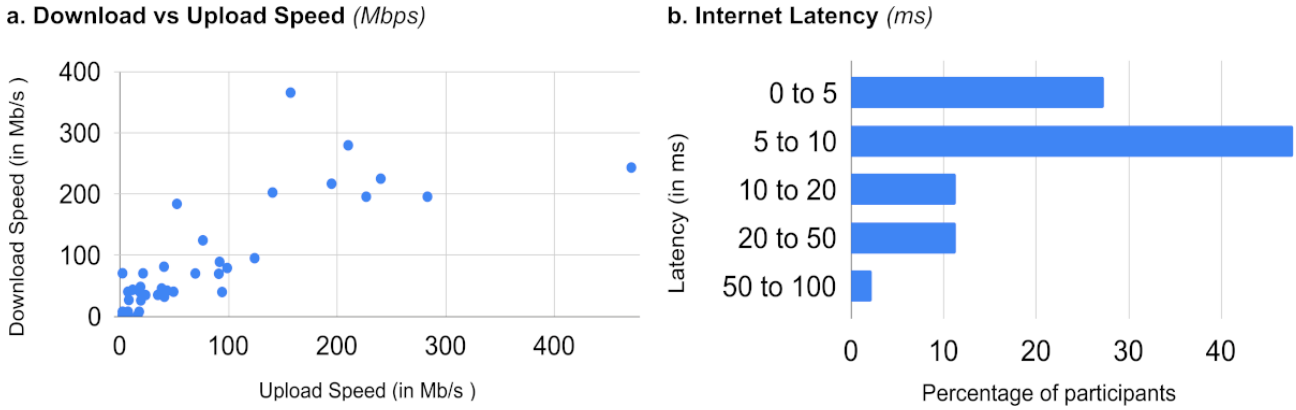


Table 4. Summary of hardware and web capability specifications for remote guided participants, compared to the standard testing equipment used for the face-to-face group.

Hardware specifications	RGT[a] n/mean, (%/SD)	F2F[b] standard
Brand		
Acer	14 (13.6%)	HP Probook
Apple	6 (31.8%)	__[c]
Asus	9 (20.5%)	—
Dell	3 (6.8%)	—
HP	7 (15.9%)	—
Lenovo	5 (11.4%)	—
Operating system		
Windows	30 (68.2%)	Windows 10
Mac OS	14 (31.8%)	—
Processor		
Intel Core i3	2 (4.5%)	Intel Core i7 2/2.4ghz
Intel Core i5	21 (47.7%)	—
Intel Core i6	1 (2.3%)	—
Intel Core i7	17 (38.6%)	—
Intel Core i8	1 (2.3%)	—
Intel Core i9	1 (2.3%)	—
Other	1 (2.3%)	—
RAM (GB)	9.73 (4.35)	8.0
Total hard disk space (GB)	417 (229)	500 HDD (+256 SSD)
Free hard disk space (GB)	270 (223)	108
Screen size (inches)	13.8 (1.74)	13.3
Screen resolution		
1280 x 800	1 (2.3%)	1920 x 1080
1366 x 768	6 (13.6%)	—
1440 x 900	2 (4.6%)	—
1920 x 1080	19 (43.2%)	—
1920 x 1280	1 (2.3%)	—
2560 x 1600	10 (22.7%)	—
3200 x 1800	3 (6.8%)	—
Unspecified	2 (4.6%)	—
Input devices		
Mouse (wireless)	27 (61.2%)	Wired mouse
Mouse (wired)	15 (34.1%)	—
Mouse (integrated)	2 (4.6%)	—
Keyboard (wireless)	2 (4.6%)	Integrated keyboard
Keyboard (integrated)	42 (95.5%)	N/A[d]
Webcam (integrated)	43 (97.7%)	Integrated webcam
Webcam (separate)	1 (2.3%)	N/A
Microphone (integrated	35 (79.6%)	Integrated microphone
Microphone (separate)	9 (20.5%)	N/A

Hardware specifications	RGT[a] n/mean, (%/SD)	F2F[b] standard
Web Capability		
download speed (Mb/s)	77.9 (88.6)	44.6
Upload speed (Mb/s)	70.4 (96.1)	48.1
Internet latency (ms)	10.6 (12.3)	5
Web browser		
Google Chrome	38 (86.4%)	Google Chrome
Mozilla Firefox	5 (2.3%)	N/A
Safari	1 (11.4%)	N/A

[a]RGT: remote guided testing.

[b]F2F: face-to-face.

[c]We used one set of standard equipment for testing the F2F participants, hence there is only one value reported for each subheading under the F2F column.

[d]N/A: not applicable.

Data Quality

Missed Trials

Table 5 shows the percentages of trials missed for each experimental task and group. To assess whether there was a difference in the number of missed trials across tasks as a function of testing modality, a general linear model (GLM) analysis with missed trials on each task (4 levels) as dependent (within-subjects) variables and modality (2 levels) as a predictor (between-subjects) variable was employed. Participants' age and vocabulary standardized scores were entered as covariates in the model. Since Mauchly's test indicated that the assumption of sphericity had been violated (χ^2_5=176; P<.001), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε=0.53). The results indicated no significant main effect of modality ($F_{1,56}$=.61; P=.44; η^2P=.01) and no significant interaction between modality and task ($F_{1.59,\ 89.3}$=.44; P=.60; η^2P=.01). Tukey HSD posthoc tests indicated that F2F and RGT participants did not differ on missed trials for any individual task (P>.99 for all pairwise comparisons). There were also no significant effects of age ($F_{1,56}$=.52; P=.47; η^2P=.01) and vocabulary ($F_{1,56}$=.01; P=.91; η^2P=.00).

Table 5. Summary of data quality indices for all tasks.

Delivery platform and task	(1) Missed trials (%), mean (SD)		(2) Data exclusion				(1) Reaction time (sec), mean (SD)	
			Trial level (%), mean (SD)		Task level (N), (tech/perf)			
	F2F[a]	RGT[b]	F2F	RGT	F2F	RGT	F2F	RGT
i-ABC								
Wisconsin Card Sort Test (WCST)	0.73 (1.3)	1.02 (1.9)	3.50 (3.2)	4.92 (5.4)	0/0	0/0	1.33 (0.18)	1.39 (0.22)
Probabilistic learning and reversal (PR)	0.30 (0.6)	0.74 (1.5)	3.06 (3.1)	5.80 (5.9)	0/1	1/1	0.90 (0.16)	1.01 (0.21)
Structure learning (SL)	3.41 (2.6)	3.27 (2.6)	0.99 (0.7)	1.72 (3.5)	0/0	1/1	1.07 (0.15)	1.04 (0.16)
Inquisit								
Color-Word Stroop	N/A[c]	N/A	3.38 (4.5)	3.28 (4.7)	0/0	1/1	0.84 (0.13)	0.87 (0.14)
Stop Signal Task (SST)	0.98 (1.8)	1.59 (3.3)	1.41 (2.4)	1.14 (2.7)	0/1	1/1	0.47 (0.08)	0.42 (0.09)
Trails A and B	0 (0)	0 (0)	0 (0)	0 (0)	0/0	2/0	40.9 (10.7)	40.4 (10.2)
CANTAB[d]								
Intra/extra-dimensional set shift (IED)	N/A	N/A	N/A	N/A	0/0	0/0	N/A	N/A
Spatial working memory (SWM)	N/A	N/A	N/A	N/A	0/0	0/0	N/A	N/A
Verbal								
Backwards Digit Span	N/A	N/A	N/A	N/A	N/A	0/0	0/0	N/A
WASI[e] vocabulary	N/A	N/A	N/A	N/A	N/A	0/0	0/0	N/A

[a]F2F: face-to-face.

[b]RGT: remote guided testing.

[c]N/A: not applicable.

[d]CANTAB: Cambridge Neuropsychological Test Automated Battery.

[e]WASI: Wechsler Abbreviated Scale of Intelligence.

Data Exclusion

Trial-Level Exclusions

Table 5 provides a full breakdown of data exclusions by experimental task and group. The data were analyzed using a GLM with excluded trials on each task (5 levels) as dependent (within-subjects) variables and modality (2 levels) as a predictor (between-subjects) variable to determine whether there was a difference in the overall percentage of excluded trials across tasks as a function of testing modality. Participants' age and vocabulary standardized scores were entered as covariates in the model. Since Mauchly's test indicated that the assumption of sphericity had been violated (χ^2_9=17.1; P=.047), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε=0.88). The results indicated no significant main effect of Modality ($F_{1,56}$=2.1; P=.15; η^2P=.04) and no significant interaction between modality and task ($F_{3.55,\ 198.5}$=.37; P=.81; η^2P=.01). Tukey HSD posthoc tests indicated that F2F and RGT participants did not differ on excluded trials for any individual task (P>.40 for all pairwise comparisons). There were also no significant effects of age ($F_{1,56}$=1.97; P=.17; η^2P=.03) and vocabulary ($F_{1,56}$=1.74; P=.19; η^2P=.03).

Task-Level Exclusions

As shown in Table 5, a total of 12 participant task-level data sets (F2F=2, RGT=10) were excluded from the analysis. Of these, 6 data sets were removed for technical reasons, and 6 were removed for performance reasons. Technical exclusions only occurred for the RGT group due to technical issues encountered during task administration (eg, OS compatibility, software/hardware issues, and environmental disruption). No data sets in the F2F group were excluded for technical reasons. For performance-related exclusions, recall that task-level data were excluded if the participant's total number of missed and outlying trials was >25% of all trials for that task. Following these criteria, 2 participant task-level data sets were removed in the F2F group, and 4 participant task-level data sets were removed in the RGT group (a total of 6 datasets removed for performance reasons).

Reaction Times

Finally, we assessed whether the mean RTs of included trials differed as a function of testing modality (see Table 5 for group

means). The data were analyzed using a GLM with RTs on each task (6 levels) as dependent (within-subjects) variables and modality (2 levels) as a predictor (between-subjects) variable. Participants' age and vocabulary standardized scores were entered as covariates in the model. Since Mauchly's test indicated that the assumption of sphericity had been violated (χ^2_{14}=176; P<.001), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (χ=0.37). The results indicated no significant main effect of modality ($F_{1,55}$=.15; P=.70; η^2P=.00) and no significant interaction between modality and task ($F_{1.86,\ 102.3}$=.21; P=.24; η^2P=.03). Tukey HSD posthoc tests indicated that F2F and RGT participants did not differ on RT for any individual task (P>.80 for all pairwise comparisons). There were also no significant effects of age ($F_{1,55}$=.00; P=.98; η^2P=.00) and vocabulary ($F_{1,55}$=.14; P=.71; η^2P=.00).

Task Performance

The task performance indices were analyzed by delivery platform. Unlike the previous data quality measures of missed/excluded trials which were computed manually using simple and uniform criteria, these performance indices varied greatly in complexity and granularity (eg, spanning estimations of strategy, accuracy/error, and timing). Most of the performance indices were also automatically calculated by the delivery software using built-in criteria and assumptions. Accordingly, we analyzed task performance separately by delivery platform to allow us to detect any testing modality differences that emerged on some platforms and their tasks, but not others. Figure 4 and Table 6 show a full breakdown of participant performance by delivery platform, task, and test modality.

Figure 4. Plot of performance indices for (a) i-ABC; (b) Inquisit; (c) CANTAB and (d) Verbally delivered tasks. Face-to-face participants are shown in dark grey bars, remote guided participants are shown in light grey bars. Error bars indicate the standard error of the mean, ***P<.001.

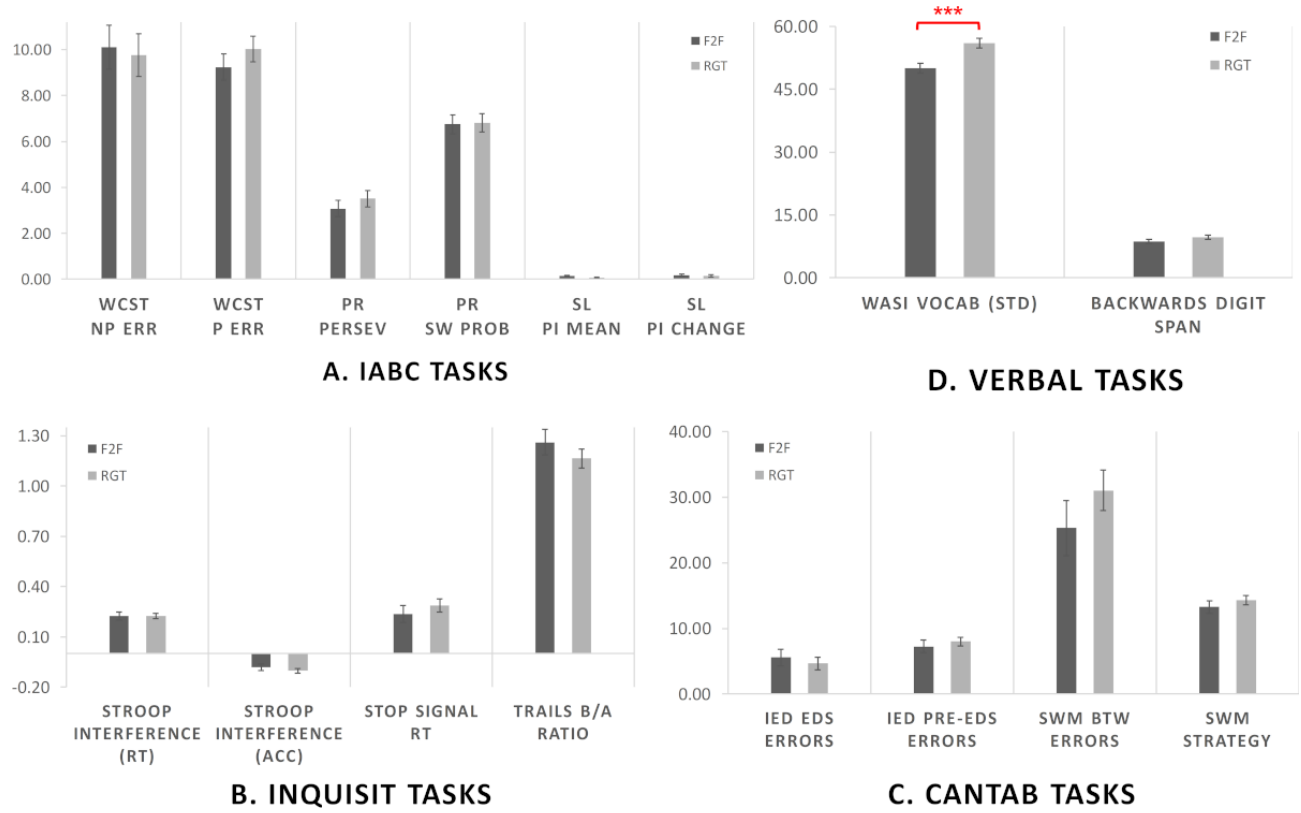


Table 6. Summary of task performance indices.

Delivery platform and task and performance index	Scores by group		GLM[a] modality effects
	F2F[b], mean (SD)	RGT[c], mean (SD)	
i-ABC			Modality $F_{1,72}$=.00; *P*=.96; Modality*Task $F_{1.20, 86.2}$=.02; *P*=.73
Wisconsin Card Sort Test (WCST)			
Nonperseverative errors	10.1 (5.5)	10.2 (6.8)	
Perseverative errors	9.3 (2.6)	10.0 (4.1)	
Probabilistic learning and reversal (PR)			
Perseveration	3.1 (1.7)	3.7 (2.8)	
Switching probability	6.6 (2.5)	6.7 (2.6)	
Structure learning (SL)			
PI mean	0.15 (0.22)	0.06 (0.18)	
PI change	0.18 (0.30)	0.16 (0.31)	
Inquisit			Modality $F_{1,58}$=.74; *P*=.39; Modality*Index $F_{1.80, 104.3}$=1.65; *P*=.20
Color-Word Stroop			
Interference (reaction time)	0.23 (0.10)	0.22 (0.11)	
Interference (accuracy)	–0.08 (0.07)	–0.10 (0.09)	
Stop Signal Task (SST)			
Stop Signal reaction time	0.24 (0.19)	0.28 (0.27)	
Trails A and B			
Trails B/A time ratio	1.26 (0.43)	1.17 (0.31)	
CANTAB[d]			Modality $F_{1,59}$=.02; *P*=.88; Modality*Task $F_{1,59}$=.03; *P*=.86
Intra-extra dimensional set shift (IED)			
Extra dimensional shift errors	5.6 (7.2)	4.7 (5.2)	
Pre-extra dimensional shift errors	7.3 (2.6)	9.6 (7.7)	
Spatial working memory (SWM)			
Between errors	25.3 (16.7)	32.3 (21.3)	
Strategy	13.3 (4.6)	14.4 (4.1)	
Verbal delivery			Modality $F_{1,82}$=16.6; *P*<.001; Modality*Task $F_{1,82}$=7.57; *P*=.007
WASI[e] vocabulary (standardized score)	50.1 (7.2)	56.0 (7.6)	
Backwards Digit Span (total score)	8.7 (3.1)	9.7 (3.2)	

[a]GLM: general linear model

[b]F2F: face-to-face.

[c]RTG: remote guided testing.

[d]CANTAB: Cambridge Neuropsychological Test Automated Battery.

[e]WASI: Wechsler Abbreviated Scale of Intelligence.

- **i-ABC:** The data were analyzed using a GLM with task (3 levels) and index (2 levels) as dependent (within-subjects) variables and testing modality (2 levels) as a predictor (between-subjects) variable to assess task performance across the three i-ABC tasks (Wisconsin Card Sort, probabilistic reversal and structure learning). Participants' age and vocabulary standardized scores were entered as covariates in the model. Since Mauchly's test indicated that the assumption of sphericity had been violated (task χ^2_2=78.7; *P*<.001), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε=0.60). The

results indicated no significant main effect of modality ($F_{1,72}$=.00; P=.96; η^2P=.00) and no significant interaction between modality and task ($F_{1.20, 86.2}$=.02; P=.92; η^2P=.00). Tukey HSD posthoc tests indicated that F2F and RGT participants did not differ on performance for any individual task or index (P>.99 for all pairwise comparisons). There was a significant effect of age ($F_{1,72}$=9.81; P=.003; η^2P=.12) but no significant effect of vocabulary ($F_{1,72}$=.02; P=.90; η^2P=.00).

- **Inquisit:** To assess task performance across the three Inquisit tasks (Stroop, Stop Signal, and Trails), the data were analyzed using a GLM with index (4 levels) as dependent (within-subjects) variables and testing modality (2 levels) as a predictor (between-subjects) variable. Participants' age and vocabulary standardized scores were entered as covariates in the model. Since Mauchly's test indicated that the assumption of sphericity had been violated (task χ^2_5=65.9; P<.001), degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity (ε=0.60). The results indicated no significant main effect of modality ($F_{1,58}$=.74; P=.39; η^2P=.01) and no significant interaction between modality and index ($F_{1.80,104.3}$=1.65; P=.20; η^2P=.03). Tukey HSD posthoc tests indicated that F2F and RGT participants did not differ on performance for any individual task or index (P>.77 for all pairwise comparisons). There were no significant effects of age ($F_{1,58}$=1.35; P=.25; η^2P=.02) or vocabulary ($F_{1,58}$=.11; P=.74; η^2P=.00).
- **CANTAB:** Performance on the two CANTAB tasks (IED shift and SWM) was analyzed using a GLM taking task (2 levels) and Index (2 levels) as dependent (within-subjects) variables and testing Modality (2 levels) as a predictor (between-subjects) variable. Participants' age and vocabulary standardized scores were entered as covariates in the model. The results indicated no significant main effect of modality ($F_{1,59}$=.02; P=.88; η^2P=.00) and no significant interaction between modality and task ($F_{1,59}$=.03; P=.86; η^2P=.00). Tukey HSD posthoc tests indicated that F2F and RGT participants did not differ on performance for any individual task or index (P>.51 for all pairwise comparisons). There was no significant effect of age ($F_{1,59}$=.32; P=.57; η^2P=.01) or vocabulary ($F_{1,59}$=1.92; P=.17; η^2P=.03).
- **Verbal delivery:** Finally, participants' performance on the verbally delivered tasks (WASI Vocabulary and Backwards Digit Span) was assessed using a GLM with task (2 levels) as dependent (within-subjects) variables and testing modality (2 levels) as a predictor (between-subjects) variable. Only participant age was entered as a covariate in the model. Unlike all the previous tests, we observed a strong and significant main effect of modality ($F_{1,82}$=16.6; P<.001; η^2P=.17) as well as a significant interaction between modality and task ($F_{1,82}$=7.57, P=.01; η^2P=.08). Tukey HSD posthoc tests of the interaction indicated that F2F and RGT participants differed significantly on vocabulary performance (P<.001, RGT>F2F) but not on digit span (P=.83). There was no significant effect of age ($F_{1,82}$=.96; P=.33; η^2P=.01).

In summary, we observed no significant difference in task performance between F2F and RGT participants for any delivery platform or experimental task, with the notable exception of WASI Vocabulary, where RGT participants scored significantly higher than F2F participants on the task.

Verbal Intelligence Analysis

To understand the source of this apparent difference in verbal intelligence, first, we assessed whether participants' background could explain their differences in vocabulary performance. Accordingly, the categorical factors of gender (2 levels, male/female), ethnicity (3 levels, Chinese/Malay/Indian), education (2 levels, secondary/bachelors), home-dwelling (6 levels), and testing modality (2 levels, F2F/RGT), and the continuous variable of age were entered as predictors in a general regression model analysis, taking vocabulary score as the dependent variable. Overall, the model was significant ($F_{11,62}$=2.89; P=.004; adjusted R^2=0.22); however, the only significant predictor of vocabulary was testing modality (β=–0.47, SE 0.11, t=–4.15; P<.001). None of the other factors (age, gender, ethnicity, education, or dwelling) significantly predicted vocabulary scores (P>.25 for all). Therefore, group differences in verbal intelligence could not be explained by differences in participant background characteristics.

Next, we conducted further analyses on participants' item-level responses. Recall that participants received 0 (for an incorrect or null response), 1 (for a partial response), or 2 points (for a full response) on each word item. We assessed whether superior performance in the RGT group was due to (1) knowledge of more words (ie, reaching a higher word item number) or (2) more complete description of words (ie, attaining a full score of 2 for a higher proportion of words). Unpaired two-tailed t tests conducted for each contrast revealed that RGT participants reached a significantly higher item number than F2F participants on average (F2F: mean 24.0, SD 1.3 and RGT: mean 25.3, SD 1.9; t_{83}=–3.79; P<.001). However, a two-tailed t test showed that RGT participants also attained a full score on a higher proportion of items than F2F participants (F2F: mean 0.53, SD 0.23 and RGT: mean 0.63, SD 0.19; t_{83}=–2.18; P=.03). Therefore, the item-level analysis supported both effects.

Discussion

Principal Findings

The COVID-19 pandemic has fundamentally changed the landscape of human psychological research and left in its wake a need for thoughtful recalibration of the balance between new remote ways of working and traditional lab-based research approaches. Never has there been greater urgency and impetus to shift toward web-based data collection methods. Yet data quality and assurance frameworks for online protocols—particularly for web-based cognitive measurements—are still lacking, and current published

web-based studies vary greatly in their data quality monitoring and transparency. Therefore, we know surprisingly little about how web-based data sets differ from data collected in person, and significant questions remain regarding experimental rigor, reliability, and validity [36,37,66]. To help identify exactly how sources of unwanted participant variability may arise during online data collection and to mitigate these effects, we propose a new supervised online testing methodology, RGT. This hybrid method may offer a close alternative to traditional lab-based methods for collecting high-quality human cognitive data without requiring physical contact in the post-COVID "new normal" where many people now work from home.

Further, although we use RGT in a research context, our findings demonstrate there is no reason that the method could not be used clinically for neuropsychological assessments, particularly in situations where in-person meetings would be difficult or impossible. For example, people in wheelchairs or care homes may find it easier to be tested in their home environment, particularly during winter when daylight hours are short, and there can be significant weather deterrents to travel (eg, ice or snow). Therefore, there is wide potential for the RGT method to be used in tandem with traditional F2F methods across both clinical and nonclinical settings.

RGT Data Quality

Three data quality indices were examined in cognitive test data collected via RGT and standard lab-based F2F methods: (1) missed trials, (2) data exclusion (both at the individual trial and participant level), and (3) RTs. The results showed that more participant data sets were excluded for technical reasons, such as hardware or software incompatibility issues, or in one case, environmental disruption in the RGT data set (n=6 across all tasks for RGT compared to none for F2F). However, RGT and F2F data sets did not differ on any of the other data quality indices of missed and excluded trials or on RT. The latter result is particularly relevant since previous web-based studies that have examined RT indices note significant and consistent lags in participant response time latencies during unsupervised web-based testing [34,76]. This indicates that experimenter supervision, even if only as a virtual presence, may be crucial for maintaining participant focus and attention on cognitive tasks, particularly when an expedient response is required. Additionally, the supervising experimenter was also able to quickly troubleshoot several common software and set-up problems that RGT participants experienced, which could have otherwise exacerbated the number of technical issues and data degradation.

It is well-established that the "experimenter effect" has a significant influence on participants' motivation, mental state, performance, task engagement, and credibility during experimental studies [87,88]. It should also be noted that since F2F testing was also supervised, experimenter effects were likely to have been similar across groups, and in this case, apparently beneficial for task compliance. However, there are scenarios in which supervision may adversely affect participants' cognitive and behavioral performance due to the social desirability effect and increased cognitive load [33,89,90]. For instance, Richman et al [89] reported decreased pressure to impress (social desirability effect) with the use of online-based settings. Therefore, the proposed RGT method may not be optimal for experimental paradigms that are sensitive to social desirability effects.

RGT Task Performance

No significant differences in task performance were observed across all measures of executive function (cognitive flexibility, working memory, and inhibition) and learning, administered using 3 different experimental platforms (CANTAB, Inquisit, and i-ABC). However, we did observe a large and unpredicted difference in verbal intelligence (vocabulary) when measured in remote and in-person settings. Surprisingly, the RGT group scored significantly higher than the F2F group, and this effect could not be explained by differences in background characteristics (age, gender, ethnicity, or socioeconomic status). Detailed analyses of the item-level responses suggested that RGT participants produced correct definitions for a significantly higher number of words and also produced more fully elaborated responses to individual test items than F2F participants. This could be due to both F2F participants and the experimenter wearing facial masks and maintaining a physical distance of at least 1m (in compliance with prevailing COVID-19 guidance) throughout the experimental session in the lab. This could have influenced participants' general willingness to communicate with the experimenter, consistent with data from a previous large-scale randomized control study indicating that mask-wearing by physicians during consultations negatively impacted doctor-patient communication, perceived empathy, and relational continuity [91]. Therefore, in clinical settings, remote testing methods not requiring the use of personal protective equipment such as masks may, in fact, be beneficial to reduce the communication barrier between experimenter and participant, thereby yielding improved performance on verbal tasks.

A strength of this study is that participants were of diverse and Asian origin (including Chinese, Malay, and Indian ethnicities), which addresses the Western skew in participant demographics that has characterized much of psychological research [42-47]. In this context, it is encouraging to note that web-based remote methodologies are suitable for these populations. However, most of the participants were highly educated university students whose attitudes, moral reasoning, beliefs, and social networks are known to differ significantly from that of nonuniversity educated counterparts [38]. Although the current study did not pertain specifically to social attitudes or phenomena, these factors may nonetheless have implicitly influenced data collection (eg, social desirability bias during experimenter monitoring, etc), limiting the broader generalizability of these findings to other populations.

Toward a Data Quality Assurance Framework for Web-Based Cognitive Studies

Given the current societal momentum, we expect to see a continued rise in the number of cognitive studies conducted using web-based protocols. There exists, therefore, an urgent need for standardized protocols, data quality assurance indices, and benchmarks for the conduct and reporting of web-based cognitive studies. We take a step in this direction by making

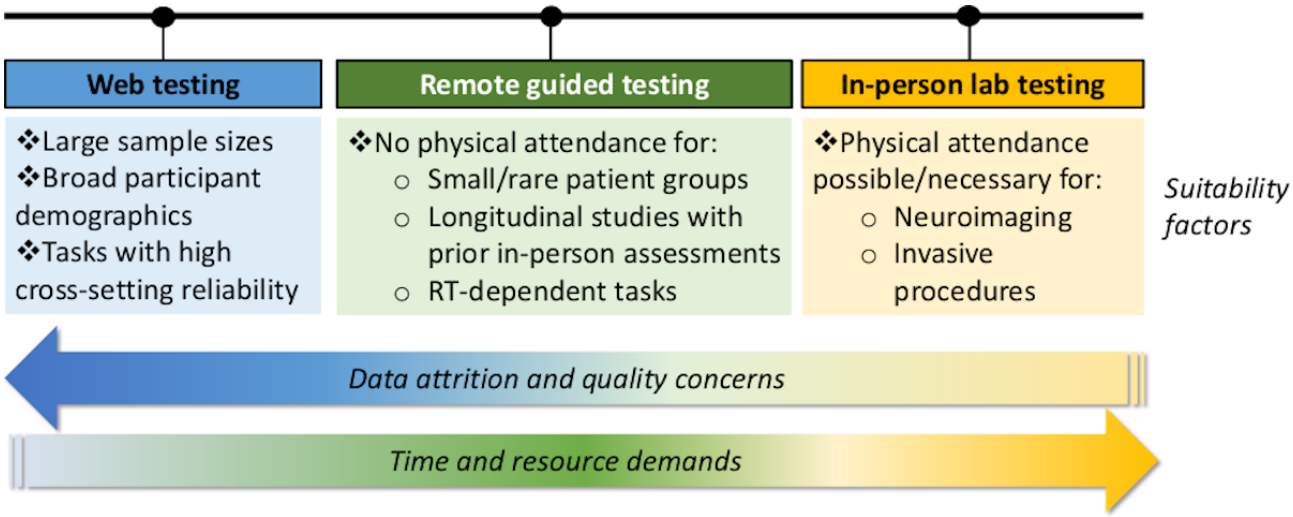
the standard operating protocol for our remote guided method freely available (Multimedia Appendix 1). We further report detailed information about participants' technological capability and home environment, including the relevant survey instruments that were developed for this purpose (Multimedia Appendices 2 and 3). We define and report a detailed set of data quality indices, which include measures of trial-level variability (eg, missed responses, outlying responses, and RTs) as well as participant variability. We further distinguish between technical-related and performance-related issues and exclusions while providing in-depth descriptions of each. This level and form of reporting may help orient the field of web-based cognitive testing toward greater transparency, reliability, and replicability and also provide common metrics on which the data quality of different datasets may be compared [36,37,66].

Limitations and Considerations for Selecting Test Methodology

Our results suggest that the RGT method yields high-quality cognitive data comparable to data collected in-person in the lab. However, this gain in data quality comes at the cost of additional manpower and time required for remote human supervision. In fact, compared to lab-based testing, the RGT method requires one additional set-up session (lasting 30 minutes) and therefore presents a greater time demand for both the experimenter and participant. This level of time investment may not be appropriate for large-scale studies that aim to test thousands of participants in a short period, although the inclusion of clear instructions for online tests and brief online tutoring (eg, using video clips) may improve the comprehensibility of instructions if the requisite research personnel are not available. As illustrated in Figure 5, one practical consideration when deciding on an appropriate test methodology is the trade-off between data quality and available time or resources. Purely unsupervised web-based testing has the attendant advantages of reaching large sample sizes with a broad demographic at a relatively low cost per head [34,49,54,56-58]. However, this may compromise data quality, comparability, replicability, and validity [61,62]. Therefore, implementing unsupervised web-based testing methods must be informed by the specific tasks to be used and their proven cross-setting reliability [34].

Another important consideration is the necessity and feasibility of in-person attendance at a physical location. Certain experimental protocols (eg, neuroimaging and invasive procedures) require in-person attendance due to the need for specific equipment or professional expertise. In these cases, in-person lab-based testing is the only option for data collection. However, in situations where physical attendance is not necessary or impossible (eg, during COVID-19 lockdown restrictions), RGT may be a viable alternative. The decision to adopt a method like RGT will be further weighted by considerations of group size and composition; for example, in clinical studies that involve high-risk or rare cohorts where maximization of individual data quality is important. Similarly, longitudinal studies that have used in-person lab-based cognitive tests at previous timepoints may prioritize cross-setting comparability, opting for supervised online methods that yield similar results to lab-based tests. Further, studies that include RT-dependent tasks (eg, Stroop and Stop Signal) may wish to use supervised online methods to ameliorate known reaction latency issues [76]. Finally, both supervised and unsupervised web-based methods require good internet connectivity and digital infrastructure for participants and research labs involved. The excellent web capability of RGT participants in the current study is indicative of Singapore having one of the highest levels of internet penetration in the world, recently estimated at 87% [92]. Therefore, while web-based testing would be highly feasible in countries like Singapore, this may be more challenging in countries with less well-developed digital infrastructure.

Figure 5. Summary of considerations for suitability of unsupervised, supervised web testing and in-person methodologies for cognitive testing. RT: reaction time.



Conclusions

The global COVID-19 pandemic has accelerated a move toward web-based cognitive testing, yet long-standing questions remain over the data quality and validity of web-based studies, compounding an urgent need to develop and implement data quality assurance frameworks for current and future online studies. Here, we propose a new supervised online testing

methodology, RGT, and present data quality benchmarks for this new method. Across all measures of data quality and performance, the RGT method yielded data that was statistically equivalent to data collected in person in the lab. We conclude that the RGT methodology is robust and offers a viable alternative for collecting high-quality human cognitive data in both lab-based research and clinical contexts without requiring in-person physical attendance.

Acknowledgments

This research was conducted by the Centre for Lifelong Learning and Individualized Cognition (CLIC). CLIC is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Campus for Research Excellence and Technological Enterprise (CREATE) program. The research was also supported by a Nanyang Technological University grant to VL (M4081585.SS0), the Ministry of Education (Singapore) Tier 1 grants to VL (M4012105.SS0 and M4011750.SS0), and grants to ZK from the Biotechnology and Biological Sciences Research Council (H012508 and BB/P021255/1), the Wellcome Trust (205067/Z/16/Z), and the European Union's Horizon 2020 research and innovation program (grant numbers 765121 and 840271).

We thank Janice Tan Yu Jin for her assistance with data collection, Avraam Papadopoulos for technical support with the i-ABC app, and Kastoori d/o Kalaivanan for assistance with proofreading.

Conflicts of Interest

TWR receives consultancy fees from Cambridge Cognition. BJS consults for Cambridge Cognition.

Multimedia Appendix 1
Standard operating protocols.
[PDF File (Adobe PDF File), 133 KB - jmir_v24i1e28368_app1.pdf]

Multimedia Appendix 2
Equipment questionnaire (remote guided testing).
[PDF File (Adobe PDF File), 64 KB - jmir_v24i1e28368_app2.pdf]

Multimedia Appendix 3
Testing environment checklist (remote guided testing).
[PDF File (Adobe PDF File), 35 KB - jmir_v24i1e28368_app3.pdf]

Multimedia Appendix 4
Task descriptions and performance indices.
[PDF File (Adobe PDF File), 450 KB - jmir_v24i1e28368_app4.pdf]

References

1. Omary M, Eswaraka J, Kimball SD, Moghe PV, Panettieri RA, Scotto KW. The COVID-19 pandemic and research shutdown: staying safe and productive. J Clin Invest 2020 Jun 01;130(6):2745-2748 [FREE Full text] [doi: 10.1172/JCI138646] [Medline: 32243259]
2. Ali SH, Foreman J, Capasso A, Jones AM, Tozan Y, DiClemente RJ. Social media as a recruitment platform for a nationwide online survey of COVID-19 knowledge, beliefs, and practices in the United States: methodology and feasibility analysis. BMC Med Res Methodol 2020 May 13;20(1):116 [FREE Full text] [doi: 10.1186/s12874-020-01011-0] [Medline: 32404050]
3. Serhan D. Transitioning from face-to-face to remote learning: Students attitudes and perceptions of using Zoom during COVID-19 pandemic. International Journal of Technology in Education and Science 2020:335-342. [doi: 10.46328/ijtes.v4i4.148]
4. Ray S, Srivastava S. Virtualization of science education: a lesson from the COVID-19 pandemic. J Proteins Proteom 2020 May 31;11(2):1-4 [FREE Full text] [doi: 10.1007/s42485-020-00038-7] [Medline: 33132627]
5. Cinelli M, Quattrociocchi W, Galeazzi A, Valensise CM, Brugnoli E, Schmidt AL, et al. The COVID-19 social media infodemic. Sci Rep 2020 Oct 06;10(1):16598 [FREE Full text] [doi: 10.1038/s41598-020-73510-5] [Medline: 33024152]
6. Rosenfeld D, Balcetis E, Bastian B, Berkman ET, Bosson JK, Brannon TN, et al. Psychological Science in the Wake of COVID-19: Social, Methodological, and Metascientific Considerations. Perspect Psychol Sci 2021 Oct 01:1745691621999374. [doi: 10.1177/1745691621999374] [Medline: 34597198]
7. Holmes EA, O'Connor RC, Perry VH, Tracey I, Wessely S, Arseneault L, et al. Multidisciplinary research priorities for the COVID-19 pandemic: a call for action for mental health science. The Lancet Psychiatry 2020 Jun;7(6):547-560 [FREE Full text] [doi: 10.1016/s2215-0366(20)30168-1]

8. Marcell M, Falls AL. Online data collection with special populations over the World Wide Web. Downs Syndr Res Pract 2001 Oct;7(3):106-123 [FREE Full text] [doi: 10.3104/reports.120] [Medline: 11721536]
9. Musch J, Reips UD. A Brief History of Web Experimenting. In: Birnbaum MH, editor. Psychological experiments on the Internet. San Diego, CA: Academic Press; 2000.
10. Krantz J, Ballard J, Scher J. Comparing the results of laboratory and World-Wide Web samples on the determinants of female attractiveness. Behavior Research Methods, Instruments, & Computers 1997 Jun;29(2):264-269 [FREE Full text] [doi: 10.3758/bf03204824]
11. Nosek BA, Banaji MR, Greenwald AG. E-Research: Ethics, Security, Design, and Control in Psychological Research on the Internet. J Social Isssues 2002 Jan;58(1):161-176 [FREE Full text] [doi: 10.1111/1540-4560.00254]
12. Zygouris S, Tsolaki M. Computerized Cognitive Testing for Older Adults: A Review. Am J Alzheimers Dis Other Demen 2015 Feb 13;30(1):13-28 [FREE Full text] [doi: 10.1177/1533317514522852] [Medline: 24526761]
13. Heaton RK, Chelune GK, Talley JL, Kay GG, Curtiss G. Wisconsin Card Sorting Test Manual: Revised and Expanded. Odessa, FL: Psychological Assessment Resources. 1993. URL: https://www.bibsonomy.org/bibtex/2db9f59c0ebfd219f28eda9773b1aa6a2/brian.mingus [accessed 2020-12-20]
14. Fortuny L, Heaton RK. Standard versus Computerized Administration of the Wisconsin Card Sorting Test. The Clinical Neuropsychologist 1996 Oct;10(4):419-424 [FREE Full text] [doi: 10.1080/13854049608406702]
15. Berg EA. A Simple Objective Technique for Measuring Flexibility in Thinking. J Gen Psychol 1948 Jul;39(1):15-22. [doi: 10.1080/00221309.1948.9918159] [Medline: 18889466]
16. Heaton RK. Wisconsin Card Sorting Test, Computer Version 4, Psychological Assessment Resources: Odessa, FL, USA. 2003. URL: https://www.acer.org/files/WCSTCV4_Report_RESEARCH.pdf [accessed 2001-12-20]
17. Kongs SK, Thompson LT, GL I. Wisconsin card sorting test - 64 card computerized version. Odessa: Psychological Assessment Resources. 2000. URL: https://www.parinc.com/Products/Pkey/479 [accessed 2020-12-03]
18. Steinke A, Kopp B, Lange F. The Wisconsin Card Sorting Test: Split-Half Reliability Estimates for a Self-Administered Computerized Variant. Brain Sci 2021 Apr 21;11(5):529 [FREE Full text] [doi: 10.3390/brainsci11050529] [Medline: 33919298]
19. Wang R, Shen Y, Tino P, Welchman AE, Kourtzi Z. Learning Predictive Statistics: Strategies and Brain Mechanisms. J Neurosci 2017 Jul 31;37(35):8412-8427 [FREE Full text] [doi: 10.1523/jneurosci.0144-17.2017]
20. MacLeod C. The Stroop task: The "Gold Standard" of Attentional Measures. Journal of Experimental Psychology: General 1992;121(1):12-14 [FREE Full text] [doi: 10.1037/0096-3445.121.1.12]
21. Logan G, Cowan WB, Davis KA. On the ability to inhibit simple and choice reaction time responses: A model and a method. Journal of Experimental Psychology: Human Perception and Performance 1984;10(2):276-291 [FREE Full text] [doi: 10.1037/0096-1523.10.2.276]
22. Verbruggen F, Aron AR, Band G, Beste C, Bissett PG, Brockett AT, et al. A consensus guide to capturing the ability to inhibit actions and impulsive behaviors in the stop-signal task. Elife 2019;8:1-26 [FREE Full text] [doi: 10.7554/eLife.46323] [Medline: 31033438]
23. Reitan RM. The Relation of the Trail Making Test to Organic Brain Damage. J Consult Psychol 1955 Oct;19(5):393-394 [FREE Full text] [doi: 10.1037/h0044509] [Medline: 13263471]
24. Kortte KB, Horner MD, Windham WK. The Trail Making Test, Part B: Cognitive Flexibility or Ability to Maintain Set? Applied Neuropsychology 2002 Jun;9(2):106-109 [FREE Full text] [doi: 10.1207/s15324826an0902_5]
25. Reynolds CR. Comprehensive Trail Making Test: Examiner's manual. Austin, Texas: PRO-E; 2002. URL: https://www.statisticssolutions.com/free-resources/directory-of-survey-instruments/comprehensive-trail-making-test-ctmt/ [accessed 2020-12-05]
26. Downes JJ, Roberts AC, Sahakian BJ, Evenden JL, Morris RG, Robbins TW. Impaired extra-dimensional shift performance in medicated and unmedicated Parkinson's disease: evidence for a specific attentional dysfunction. Neuropsychologia 1989;27(11-12):1329-1343. [doi: 10.1016/0028-3932(89)90128-0] [Medline: 2615934]
27. Owen AM, Downes JJ, Sahakian BJ, Polkey CE, Robbins TW. Planning and Spatial Working Memory following Frontal Lobe Lesions in Man. Neuropsychologia 1990 Jan;28(10):1021-1034 [FREE Full text] [doi: 10.1016/0028-3932(90)90137-d]
28. Wechsler D. Wechsler Abbreviated Scale of Intelligence - Second Edition (WASI-II). San Antonio, TX: NCS Pearson. 2011. URL: https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Abbreviated-Scale-of-Intelligence-%7C-Second-Edition/p/100000593.html [accessed 2020-12-15]
29. Wechsler D. Wechsler Adult Intelligence Scale - Fourth Edition (WAIS-IV). San Antonio, TX: NCS Pearson. 2008. URL: https://www.pearsonassessments.com/store/usassessments/en/Store/Professional-Assessments/Cognition-%26-Neuro/Wechsler-Adult-Intelligence-Scale-%7C-Fourth-Edition/p/100000392.html [accessed 2020-12-18]
30. Fink HA, Hemmy LS, MacDonald R, Carlyle MH, Olson CM, Dysken MW, et al. Cognitive Outcomes After Cardiovascular Procedures in Older Adults: A Systematic Review. Technology Assessment Report 2014:A-2. [Medline: 25905147]
31. Reynolds CR. Forward and backward memory span should not be combined for clinical analysis. Arch Clin Neuropsychol 1997;12(1):29-40. [Medline: 14588432]

32. Morrison GE, Simone CM, Ng NF, Hardy JL. Reliability and Validity of the NeuroCognitive Performance Test, a Web-based Neuropsychological Assessment. Front Psychol 2015;6:1652 [FREE Full text] [doi: 10.3389/fpsyg.2015.01652] [Medline: 26579035]
33. Wood RL, Liossi C. The Ecological Validity of Executive Tests in a Severely Brain Injured Sample. Arch Clin Neuropsychol 2006 Aug;21(5):429-437. [doi: 10.1016/j.acn.2005.06.014] [Medline: 16879945]
34. Backx R, Skirrow C, Dente P, Barnett JH, Cormack FK. Comparing Web-Based and Lab-Based Cognitive Assessment Using the Cambridge Neuropsychological Test Automated Battery: A Within-Subjects Counterbalanced Study. J Med Internet Res 2020 Aug 04;22(8):e16792 [FREE Full text] [doi: 10.2196/16792] [Medline: 32749999]
35. Schmand B. Why are Neuropsychologists so Reluctant to Embrace Modern Assessment Techniques? Clin Neuropsychol 2019 Feb;33(2):209-219 [FREE Full text] [doi: 10.1080/13854046.2018.1523468] [Medline: 30882285]
36. Bauer RM, Iverson GL, Cernich AN, Binder LM, Ruff RM, Naugle RI. Computerized Neuropsychological Assessment Devices: Joint Position Paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. Arch Clin Neuropsychol 2012 May;27(3):362-373 [FREE Full text] [doi: 10.1093/arclin/acs027] [Medline: 22382386]
37. Skitka LJ, Sargis EG. The Internet as Psychological Laboratory. Annu Rev Psychol 2006;57:529-555 [FREE Full text] [doi: 10.1146/annurev.psych.57.102904.190048] [Medline: 16318606]
38. Henrich J, Heine SJ, Norenzayan A. The weirdest people in the world? Behav Brain Sci 2010 Jun;33(2-3):61-83 [FREE Full text] [doi: 10.1017/S0140525X0999152X] [Medline: 20550733]
39. Nielsen M, Haun D, Kärtner J, Legare CH. The Persistent Sampling Bias in Developmental Psychology: A Call to Action. J Exp Child Psychol 2017 Oct;162:31-38 [FREE Full text] [doi: 10.1016/j.jecp.2017.04.017] [Medline: 28575664]
40. Arnett JJ. The Neglected 95%: Why American Psychology needs to become less American. American Psychologist 2008;63(7):602-614 [FREE Full text] [doi: 10.1037/0003-066x.63.7.602]
41. Arnett JJ. The Neglected 95%: A Challenge to Psychology's Philosophy of Science. American Psychologist 2009;64(6):571-574 [FREE Full text] [doi: 10.1037/a0016723]
42. Ardila A. Cultural values underlying psychometric cognitive testing. Neuropsychol Rev 2005 Dec 01;15(4):185-195. [doi: 10.1007/s11065-005-9180-y] [Medline: 16395623]
43. Uzzell BP, Ponton M, Ardila A, editors. International handbook of cross-cultural neuropsychology. Hillsdale, NJ: Lawrence Erlbaum Associates Inc; 2007.
44. Byrd DA, Rivera Mindt MM, Clark US, Clarke Y, Thames AD, Gammada EZ, et al. Creating an antiracist psychology by addressing professional complicity in psychological assessment. Psychol Assess 2021 Mar;33(3):279-285. [doi: 10.1037/pas0000993] [Medline: 33779204]
45. Patton DE, Duff K, Schoenberg MR, Mold J, Scott JG, Adams RL. Performance of cognitively normal African Americans on the RBANS in community dwelling older adults. Clin Neuropsychol 2003 Nov;17(4):515-530 [FREE Full text] [doi: 10.1076/clin.17.4.515.27948] [Medline: 15168916]
46. Rivera Mindt M, Byrd D, Saez P, Manly J. Increasing culturally competent neuropsychological services for ethnic minority populations: A call to action. Clin Neuropsychol 2010 Apr;24(3):429-453 [FREE Full text] [doi: 10.1080/13854040903058960] [Medline: 20373222]
47. Nell V. Cross-cultural neuropsychological assessment: Theory and practice. In: Lawrence Erlbaum Associates Inc.; Mahwah, NJ. Mahwah, NJ: Lawrence Erlbaum Associates Inc; 2000.
48. Kueider AM, Parisi JM, Gross AL, Rebok GW. Computerized cognitive training with older adults: A systematic review. PLoS One 2012;7(7):e40588 [FREE Full text] [doi: 10.1371/journal.pone.0040588] [Medline: 22792378]
49. Grootswagers T. A primer on running human behavioural experiments online. Behav Res Methods 2020 Dec;52(6):2283-2286 [FREE Full text] [doi: 10.3758/s13428-020-01395-3] [Medline: 32291730]
50. Anwyl-Irvine A, Massonnié J, Flitton A, Kirkham N, Evershed J. Gorilla in our midst: An online behavioral experiment builder. Behav Res Methods 2020 Feb 23;52(1):388-407 [FREE Full text] [doi: 10.3758/s13428-019-01237-x] [Medline: 31016684]
51. Barnett JH, Blackwell AD, Sahakian BJ, Robbins TW. The Paired Associates Learning (PAL) Test: 30 Years of CANTAB Translational Neuroscience from Laboratory to Bedside in Dementia Research. Translational Neuropsychopharmocology: Curr Top Behav Neurosci 2016;28:449-474 [FREE Full text] [doi: 10.1007/7854_2015_5001] [Medline: 27646012]
52. De Clercq A, Crombez G, Buysse A, Roeyers H. A simple and sensitive method to measure timing accuracy. Behav Res Methods Instrum Comput 2003 Feb;35(1):109-115 [FREE Full text] [doi: 10.3758/bf03195502] [Medline: 12723785]
53. Nakayama S, Tolbert TJ, Nov O, Porfiri M. Social Information as a Means to Enhance Engagement in Citizen Science - Based Telerehabilitation. Journal of the Association for Information Science and Technology 2018 Dec 23;70(6):587-595 [FREE Full text] [doi: 10.1002/asi.24147]
54. Anderson CA, Lindsay JJ, Bushman BJ. Research in the Psychological Laboratory: Truth or Triviality. Curr Dir Psychol Sci 2016 Jun 22;8(1):3-9 [FREE Full text] [doi: 10.1111/1467-8721.00002]
55. Eisenberg IW, Bissett PG, Zeynep Enkavi A, Li J, MacKinnon DP, Marsch LA, et al. Uncovering the structure of self-regulation through data-driven ontology discovery. Nat Commun 2019 May 24;10(1):2319 [FREE Full text] [doi: 10.1038/s41467-019-10301-1] [Medline: 31127115]

56. Thornton L, Batterham PJ, Fassnacht DB, Kay-Lambkin F, Calear AL, Hunt S. Recruiting for health, medical or psychosocial research using Facebook: Systematic review. Internet Interv 2016 May;4:72-81 [FREE Full text] [doi: 10.1016/j.invent.2016.02.001] [Medline: 30135792]
57. Mellis AM, Bickel WK. Mechanical Turk data collection in addiction research: Utility, concerns and best practices. Addiction 2020 Oct 24;115(10):1960-1968 [FREE Full text] [doi: 10.1111/add.15032] [Medline: 32135574]
58. Palan S, Schitter C. Prolific.ac—A subject pool for online experiments. Journal of Behavioral and Experimental Finance 2018 Mar;17:22-27. [doi: 10.1016/j.jbef.2017.12.004]
59. Koterba CH, Baum KT, Hamner T, Busch TA, Davis KC, Tlustos-Carter S, et al. COVID-19 issues related to pediatric neuropsychology and inpatient rehabilitation - challenges to usual care and solutions during the pandemic. Clin Neuropsychol 2020 Aug 26;34(7-8):1380-1394. [doi: 10.1080/13854046.2020.1811892] [Medline: 32847476]
60. Singh S, Germine L. Technology meets tradition: A hybrid model for implementing digital tools in neuropsychology. Int Rev Psychiatry 2021 Jun 25;33(4):382-393. [doi: 10.1080/09540261.2020.1835839] [Medline: 33236657]
61. Adjerid I, Kelley K. Big data in psychology: A framework for research advancement. Am Psychol 2018 Oct;73(7):899-917 [FREE Full text] [doi: 10.1037/amp0000190] [Medline: 29469579]
62. Casler K, Bickel L, Hackett E. Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. Computers in Human Behavior 2013 Nov;29(6):2156-2160 [FREE Full text] [doi: 10.1016/j.chb.2013.05.009]
63. Germine LT, Duchaine B, Nakayama K. Where cognitive development and aging meet: Face learning ability peaks after age 30. Cognition 2011 Feb;118(2):201-210. [doi: 10.1016/j.cognition.2010.11.002] [Medline: 21130422]
64. Germine LT, Hooker CI. Face emotion recognition is related to individual differences in psychosis-proneness. Psychol Med 2011 May;41(5):937-947. [doi: 10.1017/S0033291710001571] [Medline: 20810004]
65. Germine L, Nakayama K, Duchaine BC, Chabris CF, Chatterjee G, Wilmer JB. Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. Psychon Bull Rev 2012 Oct 25;19(5):847-857. [doi: 10.3758/s13423-012-0296-9] [Medline: 22829343]
66. Bridges D, Pitiot A, MacAskill MR, Peirce JW. The timing mega-study: Comparing a range of experiment generators, both lab-based and online. PeerJ 2020;8:e9414 [FREE Full text] [doi: 10.7717/peerj.9414] [Medline: 33005482]
67. Hansen TI, Lehn H, Evensmoen HR, Håberg A. Initial assessment of reliability of a self-administered web-based neuropsychological test battery. Computers in Human Behavior 2016 Oct;63:91-97 [FREE Full text] [doi: 10.1016/j.chb.2016.05.025]
68. Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological research online: report of Board of Scientific Affairs' Advisory Group on the Conduct of Research on the Internet. Am Psychol 2004;59(2):105-117. [doi: 10.1037/0003-066X.59.2.105] [Medline: 14992637]
69. Williams KD, Cheung CKT, Choi W. Cyberostracism: effects of being ignored over the Internet. J Pers Soc Psychol 2000 Nov;79(5):748-762. [doi: 10.1037//0022-3514.79.5.748] [Medline: 11079239]
70. Williams KD, Govan CL, Croker V, Tynan D, Cruickshank M, Lam A. Investigations into differences between social- and cyberostracism. Group Dynamics: Theory, Research, and Practice 2002;6(1):65-77. [doi: 10.1037/1089-2699.6.1.65]
71. Wang RY, Strong DM. Beyond Accuracy: What Data Quality Means to Data Consumers. Journal of Management Information Systems 1996;12(4):5-34 [FREE Full text] [doi: 10.1080/07421222.1996.11518099]
72. Feenstra HEM, Vermeulen IE, Murre JMJ, Schagen SB. Online cognition: factors facilitating reliable online neuropsychological test results. Clin Neuropsychol 2016;31(1):59-84 [FREE Full text] [doi: 10.1080/13854046.2016.1190405] [Medline: 27266677]
73. Andrews D, Nonnecke B, Preece J. Electronic Survey Methodology: A Case Study in Reaching Hard-to-Involve Internet Users. International Journal of Human-Computer Interaction 2003;16(2):185-210 [FREE Full text] [doi: 10.1207/s15327590ijhc1602_04]
74. Mulgund P, Sharman R, Anand P, Shekhar S, Karadi P. Data Quality Issues With Physician-Rating Websites: Systematic Review. J Med Internet Res 2020;22(9):e15916 [FREE Full text] [doi: 10.2196/15916] [Medline: 32986000]
75. Zhou H, Fishbach A. The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. J Pers Soc Psychol 2016;111(4):493-504 [FREE Full text] [doi: 10.1037/pspa0000056] [Medline: 27295328]
76. Schubert TW, Murteira C, Collins EC, Lopes D. ScriptingRT: A Software Library for Collecting Response Latencies in Online Studies of Cognition. PLoS One 2013;8(6):e67769 [FREE Full text] [doi: 10.1371/journal.pone.0067769] [Medline: 23805326]
77. Kim J, Gabriel U, Gygax P. Testing the effectiveness of the Internet-based instrument PsyToolkit: A comparison between web-based (PsyToolkit) and lab-based (E-Prime 3.0) measurements of response choice and response time in a complex psycholinguistic task. PLoS One 2019;14(9):e0221802 [FREE Full text] [doi: 10.1371/journal.pone.0221802] [Medline: 31483826]
78. Baniqued PL, Lee H, Voss MW, Basak C, Cosman JD, Desouza S, et al. Selling points: What cognitive abilities are tapped by casual video games? Acta Psychol (Amst) 2013;142(1):74-86 [FREE Full text] [doi: 10.1016/j.actpsy.2012.11.009] [Medline: 23246789]

79. Hicks KL, Foster JL, Engle RW. Measuring Working Memory Capacity on the Web with the Online Working Memory Lab (the OWL). Journal of Applied Research in Memory and Cognition 2016;5(4):478-489 [FREE Full text] [doi: 10.1016/j.jarmac.2016.07.010]
80. Ruiz S, Chen X, Rebuschat P, Meurers D. Measuring individual differences in cognitive abilities in the lab and on the web. PLoS One 2019;14(12):e0226217 [FREE Full text] [doi: 10.1371/journal.pone.0226217] [Medline: 31825980]
81. Buitenweg JIV, van de Ven RM, Prinssen S, Murre JMJ, Ridderinkhof KR. Cognitive Flexibility Training: A Large-Scale Multimodal Adaptive Active-Control Intervention Study in Healthy Older Adults. Front Hum Neurosci 2017;11:529 [FREE Full text] [doi: 10.3389/fnhum.2017.00529] [Medline: 29209183]
82. Von Gunten CD, Bartholow BD, Martins JS. Inhibition tasks are not associated with a variety of behaviors in college students. Eur J Pers 2020;34(3):412-430 [FREE Full text] [doi: 10.1002/per.2250] [Medline: 33250572]
83. Karlaftis VM, Giorgio J, Vértes PE, Wang R, Shen Y, Tino P, et al. Multimodal imaging of brain connectivity reveals predictors of individual decision strategy in statistical learning. Nat Hum Behav 2019;3(3):297-307 [FREE Full text] [doi: 10.1038/s41562-018-0503-4] [Medline: 30873437]
84. Swainson R, Rogers RD, Sahakian BJ, Summers BA, Polkey CE, Robbins TW. Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: possible adverse effects of dopaminergic medication. Neuropsychologia 2000;38(5):596-612 [FREE Full text] [doi: 10.1016/s0028-3932(99)00103-7]
85. Draine SC. Inquisit 5. Seattle: Millisecond Software. 2016. URL: https://www.millisecond.com [accessed 2020-12-10]
86. Robbins TW, James M, Owen AM, Sahakian BJ, Lawrence AD, McInnes L, et al. A study of performance on tests from the CANTAB battery sensitive to frontal lobe dysfunction in a large sample of normal volunteers: Implications for theories of executive functioning and cognitive aging. J Int Neuropsychol Soc 1998;4(5):474-490. [doi: 10.1017/s1355617798455073] [Medline: 9745237]
87. Rosenthal R. Experimenter effects in behavioral research. Irvington Publishers 1976.
88. Hoskins LL, Binder LM, Chaytor NS, Williamson DJ, Drane DL. Comparison of oral and computerized versions of the word memory test. Arch Clin Neuropsychol 2010;25(7):591-600 [FREE Full text] [doi: 10.1093/arclin/acq060] [Medline: 20719808]
89. Richman WL, Kiesler S, Weisband S, Drasgow F. A meta-analytic study of social desirability distortion in computer-administered questionnaires, traditional questionnaires, and interviews. Journal of Applied Psychology 1999;84(5):754-775 [FREE Full text] [doi: 10.1037/0021-9010.84.5.754]
90. Joinson A. Social desirability, anonymity, and Internet-based questionnaires. Behav Res Methods Instrum Comput 1999;31(3):433-438 [FREE Full text] [doi: 10.3758/bf03200723] [Medline: 10502866]
91. Wong CKM, Yip BHK, Mercer S, Griffiths S, Kung K, Wong MCS, et al. Effect of facemasks on empathy and relational continuity: A randomised controlled trial in primary care. BMC Fam Pract 2013;14:200 [FREE Full text] [doi: 10.1186/1471-2296-14-200] [Medline: 24364989]
92. Infocomm Media Development Authority. 2019. URL: https://www.imda.gov.sg/industry-development/facts-and-figures/infocomm-usage-households-and-individuals [accessed 2020-10-16]

Abbreviations

CANTAB: Cambridge Neuropsychological Test Automated Battery
CLIC: Centre for Lifelong Learning and Individualized Cognition
F2F: face-to-face
GLM: general linear model
HN: hybrid neuropsychology
IED: intra-extra dimensional
RGT: remote guided testing
RT: reaction time
SWM: spatial working memory
WASI: Wechsler Abbreviated Scale of Intelligence

Edited by R Kukafka; submitted 06.03.21; peer-reviewed by S Singh, S Lalmuanawma; comments to author 10.05.21; revised version received 25.06.21; accepted 27.07.21; published 06.01.22.

Please cite as:
Leong V, Raheel K, Sim JY, Kacker K, Karlaftis VM, Vassiliu C, Kalaivanan K, Chen SHA, Robbins TW, Sahakian BJ, Kourtzi Z
A New Remote Guided Method for Supervised Web-Based Cognitive Testing to Ensure High-Quality Data: Development and Usability Study
J Med Internet Res 2022;24(1):e28368
URL: https://www.jmir.org/2022/1/e28368
doi:10.2196/28368
PMID:34989691

Original Paper

Accuracy Assessment of Oura Ring Nocturnal Heart Rate and Heart Rate Variability in Comparison With Electrocardiography in Time and Frequency Domains: Comprehensive Analysis

Rui Cao[1], BSc; Iman Azimi[2], PhD; Fatemeh Sarhaddi[2], MSc; Hannakaisa Niela-Vilen[3], PhD; Anna Axelin[3], PhD; Pasi Liljeberg[2], PhD; Amir M Rahmani[1,4,5], MBA, PhD

[1]Department of Electrical Engineering and Computer Science, University of California, Irvine, CA, United States

[2]Department of Computing, University of Turku, Turku, Finland

[3]Department of Nursing Science, University of Turku, Turku, Finland

[4]Department of Computer Science, University of California, Irvine, CA, United States

[5]School of Nursing, University of California, Irvine, CA, United States

Corresponding Author:
Rui Cao, BSc
Department of Electrical Engineering and Computer Science
University of California
1407 Palo Verde Rd
Irvine, CA, 92617
United States
Phone: 1 6266883017
Email: caor6@uci.edu

Abstract

Background: Photoplethysmography is a noninvasive and low-cost method to remotely and continuously track vital signs. The Oura Ring is a compact photoplethysmography-based smart ring, which has recently drawn attention to remote health monitoring and wellness applications. The ring is used to acquire nocturnal heart rate (HR) and HR variability (HRV) parameters ubiquitously. However, these parameters are highly susceptible to motion artifacts and environmental noise. Therefore, a validity assessment of the parameters is required in everyday settings.

Objective: This study aims to evaluate the accuracy of HR and time domain and frequency domain HRV parameters collected by the Oura Ring against a medical grade chest electrocardiogram monitor.

Methods: We conducted overnight home-based monitoring using an Oura Ring and a Shimmer3 electrocardiogram device. The nocturnal HR and HRV parameters of 35 healthy individuals were collected and assessed. We evaluated the parameters within 2 tests, that is, values collected from 5-minute recordings (ie, short-term HRV analysis) and the average values per night sleep. A linear regression method, the Pearson correlation coefficient, and the Bland–Altman plot were used to compare the measurements of the 2 devices.

Results: Our findings showed low mean biases of the HR and HRV parameters collected by the Oura Ring in both the 5-minute and average-per-night tests. In the 5-minute test, the error variances of the parameters were different. The parameters provided by the Oura Ring dashboard (ie, HR and root mean square of successive differences [RMSSD]) showed relatively low error variance compared with the HRV parameters extracted from the normal interbeat interval signals. The Pearson correlation coefficient tests (P<.001) indicated that HR, RMSSD, average of normal heart beat intervals (AVNN), and percentage of successive normal beat-to-beat intervals that differ by more than 50 ms (pNN50) had high positive correlations with the baseline values; SD of normal beat-to-beat intervals (SDNN) and high frequency (HF) had moderate positive correlations, and low frequency (LF) and LF:HF ratio had low positive correlations. The HR, RMSSD, AVNN, and pNN50 had narrow 95% CIs; however, SDNN, LF, HF, and LF:HF ratio had relatively wider 95% CIs. In contrast, the average-per-night test showed that the HR, RMSSD, SDNN, AVNN, pNN50, LF, and HF had high positive relationships (P<.001), and the LF:HF ratio had a moderate positive relationship (P<.001). The average-per-night test also indicated considerably lower error variances than the 5-minute test for the parameters.

Conclusions: The Oura Ring could accurately measure nocturnal HR and RMSSD in both the 5-minute and average-per-night tests. It provided acceptable nocturnal AVNN, pNN50, HF, and SDNN accuracy in the average-per-night test but not in the 5-minute test. In contrast, the LF and LF:HF ratio of the ring had high error rates in both tests.

(J Med Internet Res 2022;24(1):e27487) doi:10.2196/27487

KEYWORDS

electrocardiography; ECG; wearable device; heart rate variability; Oura smart ring

Introduction

Background

Wearable devices are widely used for continuous monitoring of health parameters, by which individuals' health and well-being can be assessed [1,2]. Heart rate (HR) and HR variability (HRV) are essential parameters that can be collected noninvasively, indicating information about the cardiorespiratory and autonomic nervous systems. HRV is the variation in the time interval between adjacent heartbeats, also known as interbeat interval (IBI) [3]. Using the IBI, various parameters can be extracted, such as the root mean square of successive differences (RMSSD), SD of beat-to-beat intervals (SDNN), and the percentage of successive beat-to-beat intervals that differ by more than 50 ms (pNN50), each of which reveals various cardiovascular events and problems [4]. For example, HRV parameters have been shown to be predictors of mortality after myocardial infarction [5] and the mode of death in chronic heart failure [6]. Studies also indicated that HRV parameters are associated with diabetes [7], cardiovascular autonomic imbalance [8], and in pregnant women with pre-eclampsia [9], to mention a few. Moreover, HRV parameters are significantly correlated with sleep stage [10], sleep quality [11], and stress levels [12,13].

HR and HRV monitoring can be performed by leveraging noninvasive and low-cost methods. Electrocardiography (ECG) is a conventional method to record the heart's electrical activities, using electrodes attached to the chest and limbs [14]. ECG is a gold standard method for collecting heartbeats and IBI, as the collected electrical signals can clearly indicate depolarization of the ventricular muscles (ie, R-peak). However, the ECG method cannot be used for long-term or remote health monitoring owing to the complicated setup as ECG electrodes need to be attached to the user's limbs or chest all the time. Loose or misplaced electrode connections in the monitoring also negatively affect signal quality. Photoplethysmography (PPG) is another technique used to collect HR and HRV [15]. Different studies have focused on monitoring and extraction of PPG signals [16-18]. As an optical technique, PPG measures cyclical oscillations in the skin's blood flow by emitting light to the skin and absorbing light reflection via a light detector [19]. The light emitter and detector can be placed on the user's wrist or finger for data collection. PPG is easy to implement in remote health monitoring systems, and it is already available in various wearable devices in the market, such as smartwatches and rings.

Several clinical and commercial PPG-based smart wearable devices have been proposed in the past few years, enabling the monitoring of vital signs outside conventional clinical settings. Studies have exploited wearables such as Garmin, Fitbit, and Apple watches in clinical trials as well as in different population-based studies [20-22]. The use of wearable devices is expected to increase even further as they become smaller, lighter, and more energy-efficient with sufficient battery capacity and internal data storage. In particular, smart rings such as the Oura Ring [23] have recently drawn attention for use in remote health and physical activity monitoring, such as COVID-19 research at the University of California San Francisco [24] and players' health data monitoring in the National Basketball Association and Women's National Basketball Association league [25].

Wearable devices require a high level of accuracy and reliability, particularly if they are used in health monitoring apps. However, these devices are susceptible to artifacts, resulting in poor data collection and subsequently invalid health parameters. This problem is further exacerbated by environmental noise and motion artifacts in PPG-based monitoring [19,26]. Unfortunately, such artifacts are inevitable in free-living conditions, as the user might engage in various physical activities in different environments.

The accuracy of different HRV parameters depends on multiple factors in the signals. For example, RMSSD shows short-term variations in the IBI signal, and the accuracy is affected if a small portion of the signal is distorted [4]. In contrast, SDNN indicates the long-term signal variation, so the outliers, affecting the variation of the signal, would negatively impact its accuracy. Moreover, frequency domain features are significant for assessing the cardiovascular and nervous systems, for example, low frequency (LF) and high frequency (HF) are indicators of stress states, hypertension, and Parkinson disease severity [27,28]. These features indicate the power of the IBI in specific frequency bands. Therefore, they are distorted if interference with the same frequency is added to the signal. Such different characteristics of HRV parameters necessitate the evaluation of HRV parameters separately. Consequently, a more extensive assessment is required to investigate HRV measurements in remote monitoring.

Various studies in the literature have investigated the validity of wristbands—such as Apple Watch, Huawei Watch, and Microsoft Band 2—in terms of the quality of PPG and HRV measurements [29-32]. However, the validation of smart rings, which use finger-based PPG, is limited. Mehrabadi et al [33] assessed the nonstaging sleep parameters collected by the Oura Ring in comparison with a medically approved actigraphy device. They showed that the sleep parameters of the ring were significantly correlated with those obtained from the Actigraph. Kinnunen et al [34,35] investigated the Oura Ring via overnight data collection. The study showed good agreement between the

Oura Ring and the ECG monitoring device. However, the assessments were restricted to the nocturnal HR and RMSSD reported by the ring. Other parameters, such as the IBI or frequency domain HRV parameters, were not considered.

Objectives

In this study, we have comprehensively assessed the validity of the Oura Ring in terms of HR and multiple HRV parameters during sleep. The ring was evaluated against a medical grade chest ECG monitor. The study, approved by the ethical committee, included overnight home-based monitoring of 35 healthy individuals from whom the HR and IBI values were collected. We extracted HR, RMSSD, average of all normal heart beat intervals (AVNN), SDNN, pNN50, LF, HF, and LF:HF ratio from the ring and ECG monitor. Then, we evaluated the parameters obtained from the 2 devices in a 5-minute test and an average-per-night test. The parameters were compared using a linear regression method, Pearson correlation coefficient, and Bland–Altman plot. Finally, we have discussed the obtained results, the validity of monitoring these parameters in everyday settings, and the limitations of the study. In summary, the main contributions of this study are as follows:

1. We investigated the validity of the Oura Ring in terms of nocturnal HR and multiple HRV, compared with a medical grade chest ECG monitor.
2. We conducted a 1-day study where 35 healthy individuals were monitored at home.
3. We analyzed the HR and HRV parameters in 5-minute and average-per-night tests using a linear regression method, Pearson correlation coefficient, and Bland–Altman plot.

Methods

Study Design

The assessment of HR and HRV measurements collected from the Oura Ring was performed in an observational study in free-living conditions with a convenience sample of healthy individuals. The measurements were evaluated in comparison with a gold standard ECG monitor. Recruitment took place during July and August 2019 in southwest Finland.

Participants and Recruitment

A total of 46 healthy volunteer adults—including 23 women and 23 men—were recruited in this study. The exclusion criteria were as follows: (1) diagnosed cardiovascular disease, (2) symptoms of illness during the recruitment time, (3) restriction in physical activity, and (4) restriction on using wearable devices. The average age and BMI of the selected participants were 32.3 (SD 6.4) years and 24.9 (SD 4.5) kg/m^2, respectively. In this setup, we focused on healthy people to evaluate the accuracy of the ring, as diseases (arrhythmias) alter the shape of the PPG signal [36] and affect the regular accuracy of the ring.

In face-to-face meetings, the participants were informed about the study's detailed information, including the purpose of the study and use of wearable devices. The participants were asked to wear an Oura Ring and a Shimmer3 ECG monitor for 1 day. Measurements were conducted during normal life. A total of 11 participants were excluded from the data analysis owing to technical and practical issues, for example, the ECG electrodes were not adequately attached to the skin during sleep. Consequently, data from 35 participants (women: 19/35, 54%; men: 16/35, 46%) were included in the analysis.

Data Collection

The home-based data collection was performed using 2 wearable devices, that is, the Oura Ring [23] and Shimmer device [37]. The participants were asked to wear 1 Oura Ring on 1 finger of the nondominant hand. The Shimmer unit was placed on the chest of each participant using a chest strap. A total of 4 electrodes were attached to collect 3 bipolar limb leads during the monitoring. More details of the setup can be found in the study by Burns et al [38]. Moreover, the participants were asked to complete a short background questionnaire before starting the monitoring. They were also asked to report events during the study, for example, if the devices were removed from the finger or chest. In addition to the verbal instructions, the participants received written guidelines for using the devices.

Oura Ring is a commercial wearable device, collecting PPG, acceleration, and body temperature data to measure HR, respiratory rate, HRV, sleep parameters, and intensity of physical activity. The ring is small (2.55 mm thickness), light (4-6 g), and easy to use for continuous monitoring [35]. Its battery can support 5-7 consecutive days of monitoring with one battery charge. The ring uses Bluetooth to send data to the Oura Android or iOS operating system mobile app and cloud server. The data can be accessed through the mobile app or the server. In this study, we extracted the data from the Oura cloud [35].

The Shimmer3 ECG is the baseline device selected in this study to evaluate the HR and HRV of the ring. The device is light (31 g) and has compact dimensions (65 mm×32 mm×12 mm) [37]. The Shimmer3 ECG unit can be configured to measure ECG, accelerometer, and gyroscope data continuously. The device has a sufficient battery life and internal memory to perform monitoring for an entire day. We selected 512 Hz as the sampling frequency for ECG data collection [39,40]. The sampling frequency is sufficient to accurately extract HR and HRV [41]. Data were extracted from the device after the monitoring [37].

Data Analysis

Oura Ring

The Oura Ring provides various parameters regarding the health, physical activity, and sleep of the user. PPG-based wearable devices (including Oura) collect noise along with the signals of interest, particularly in home-based monitoring. In this study, we assessed the accuracy level or noise level of HR and HRV measurements. The ring provides HR and RMSSD when the user is sleeping, and the values are reported every 5 minutes. More details about the HR and RMSSD calculation can be found in the study by Shaffer et al [4]. The ring also provides an IBI signal [42]. We used the 5-minute window of the IBI signal to calculate the time domain parameters (ie, AVNN, SDNN, and pNN50) and frequency domain parameters (ie, LF, HF, and LF:HF ratio). The HRV parameters are presented in Table 1. It

should be noted that the ring preprocessed the signals and provided confidence values, demonstrating the validity of the IBI signals. We calculated the HRV of the 5-minute IBI signals, if at least 30% of the signal is valid [35].

Table 1. Heart rate variability parameters.

Parameter	Units	Description
NN[a] interval	ms	Time interval between 2 successive normal heartbeats
RMSSD[b]	ms	The RMSSD between adjacent NN intervals
AVNN[c]	ms	Average value of NN intervals
SDNN[d]	ms	SD of NN intervals
pNN50[e]	—[f]	The proportion of number of pairs of successive NN intervals differing more than 50 ms divided by total number of NN intervals
LF[g]	ms^2	Power of the LF band of the IBI[h] signal (ie, 0.04-0.15 Hz)
HF[i]	ms^2	Power of the HF band of the IBI signal (ie, 0.15-0.4 Hz)
LF:HF	—	Ratio of LF to HF

[a]NN: normal heart beat.

[b]RMSSD: root mean square of successive differences between normal heartbeats.

[c]AVNN: average of normal heartbeat intervals.

[d]SDNN: SD of normal beat-to-beat intervals.

[e]pNN50: percentage of successive beat-to-beat intervals that differ by more than 50 ms.

[f]Not available.

[g]LF: low frequency.

[h]IBI: interbeat interval.

[i]HF: high frequency.

Shimmer3 ECG

As previously mentioned, ECG was selected as the gold standard method to extract HR and HRV. In this regard, we chose Lead II (right arm–left leg) to extract the cardiac rhythm accurately. As the Oura Ring data were reported every 5 minutes, we also divided the ECG signals into 5-minute time windows. Then, we performed an ECG analysis to calculate the HR and HRV parameters for each window. Different steps of the analysis are illustrated in Figure 1.

The collected ECG signal quality is susceptible to artifacts generated by the user's movements, poor electrode contact, or environmental noise. Such artifacts are inevitable in home-based monitoring, as users might engage in various physical activities. In this regard, we first used a Butterworth band-pass filter with 0.5-100 Hz cutoff frequencies to remove the artifacts that were not in the desired frequency range.

We designed a two-round peak detection method to extract the R peaks from the ECG signals. In the first round, the algorithm computes the average value of the ECG in a 5-minute window. It then detects all possible peaks, including the real R peaks and miscalculated peaks (made by P wave, T wave, or noises) using the average value as the threshold. In the second round, the algorithm calculates the average value of the peaks detected in the first round. Using the average value and the normal frequency (50-200 beats per minute) of heart beats, undetected R peaks were added and miscalculated peaks were removed. Our peak detection method obtains higher accuracy than the Pan–Tompkins [43] and Hamilton [44] algorithms. Figure 2 shows the peak detection results of a sample with a 5-minute ECG window.

However, our peak detection algorithm is inaccurate if the collected ECG includes too much noise (ie, low signal-to-noise ratio). To avoid such inaccurate peak detection, we developed a method to detect and remove the distorted signals and invalid peaks. The removal criteria are based on the normal range of the HR and RR intervals learned and modified from the ECG signal quality index [45]. Such a method is important in the analysis to prevent false peak detection and, subsequently, HR and HRV extraction. The method pipeline is illustrated in Figure 3.

Figure 1. The electrocardiography analysis steps. ECG: electrocardiography; HRV: heart rate variability.

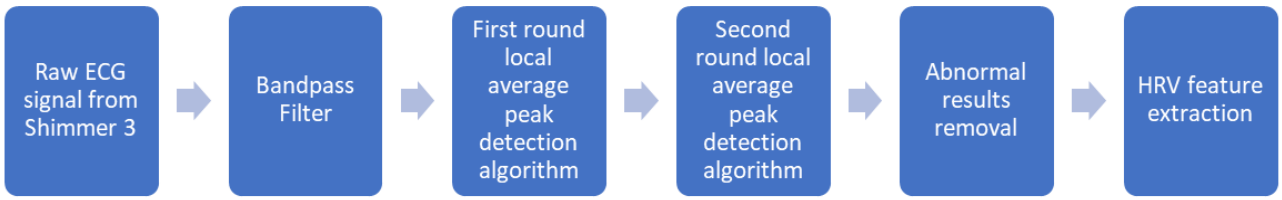


Figure 2. The peak detection result of a 5-minute time window. ECG: electrocardiography.

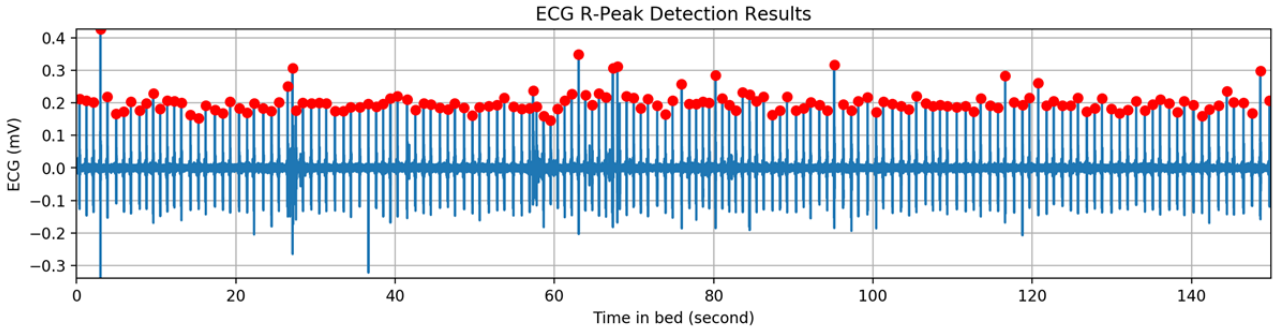
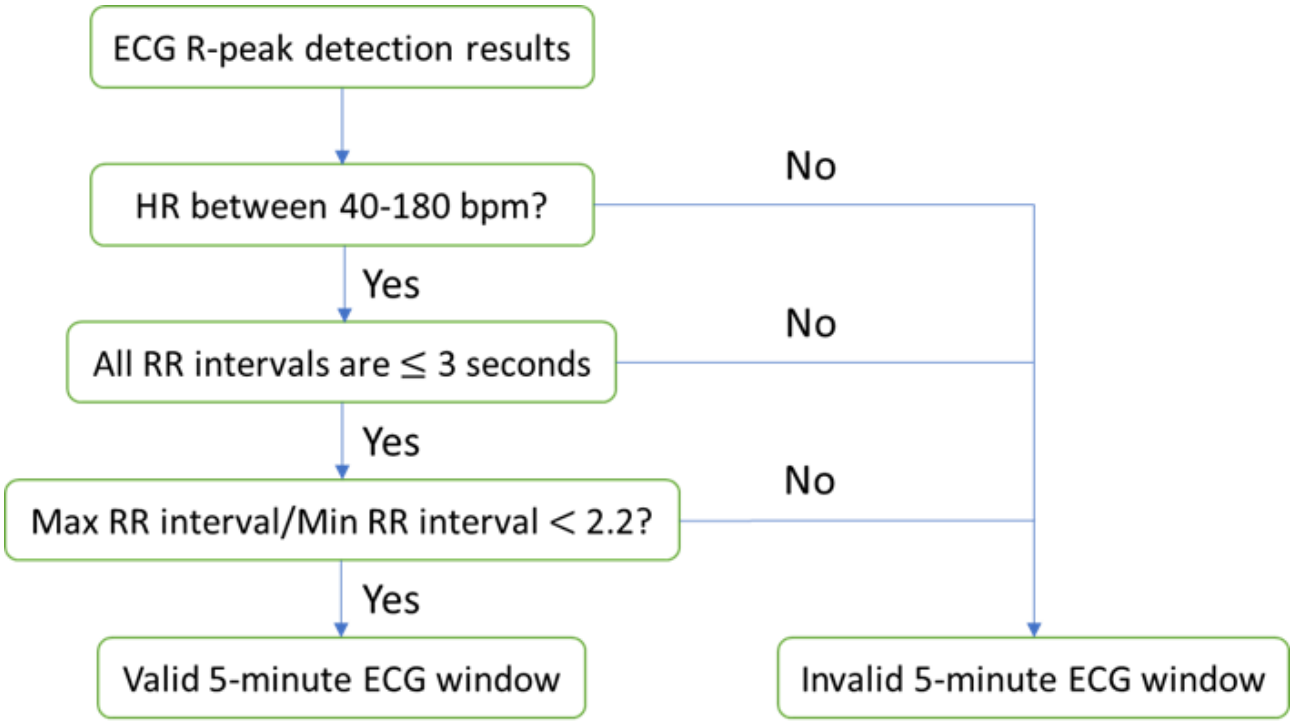


Figure 3. The pipeline of the electrocardiography validation method. ECG: electrocardiography; HR: heart rate; RR: respiratory rate.



Statistical Analysis

We used the Pearson correlation coefficient on pairwise HR and HRV parameters to investigate the linear relationship and comparability between the 2 devices. Moreover, a linear regression analysis was used to assess the accuracy of the HR and HRV parameters of the Oura Ring. We used Oura's data points (HR and HRV parameters) to fit the linear regression line. Then, we computed the R-squared value (r^2) using the regression line and corresponding baseline data points from the ECG to evaluate the closeness of the baseline data to the Oura Ring's fitted regression line. Finally, the Bland–Altman analysis was used to illustrate and estimate the agreement between the 2 devices. This method provides mean bias, SD, and 95% CIs based on the differences between the ring and Shimmer3. We used Python and Python libraries, including Scipy [46], SKlearn [47], and Statsmodels [48] to program the statistical analysis functions.

Research Ethics

The study was conducted according to the ethical principles of the Declaration of Helsinki and the Finnish Medical Research Act (No 488/1999). The study protocol was approved by the ethics committee (University of Turku, Ethics Committee for

Human Sciences, Statement no: 44/2019). The participants were informed about the study, both orally and in writing, before written informed consent was obtained. Participation was voluntary, and all participants had the right to withdraw from the study at any time and without giving any reason. To compensate for the time used for the study, each participant received a gift card to the grocery store (€20; US $26.83) at the end of the monitoring period when returning the devices.

Results

Overview

The data of 35 participants (ie, 19 women and 16 men) were included in the analysis. In this study, an average 8.25 (SD 1.51) hours of nighttime sleep data were recorded for each participant to validate HR and HRV parameters. In the following, we first evaluated the HR and HRV parameters obtained from 5-minute segments. We then compared the average parameters during nighttime sleep.

Comparisons of HR and HRV Parameters of Ring and Shimmer3 in 5-Minute Time Windows

We first investigated the correlation between HR and HRV parameters of the Oura Ring and Shimmer3 in 5-minute windows. The Pearson correlation coefficient and corresponding *P* values for the HR and HRV parameters are shown in Table 2. The HR and RMSSD between the Oura Ring and ECG were significantly correlated at $P<.001$. There were high positive relationships in the AVNN and pNN50 values, moderate positive relationships in the SDNN and HF values, and a low positive relationship in the LF and LF:HF ratio.

We used regression analysis to examine the accuracy of the Oura Ring data compared with the ECG. The regression lines (in red) for 5-minute samples of all participants are illustrated in Figure 4. We also showed y=x lines (in black), indicating the best scenario where the values obtained from the Oura and ECG are equal. Moreover, r^2 values were reported, showing the scatter of the data around the regression lines. In this analysis, the fitted lines of the HR, RMSSD, AVNN, and pNN50 were close to the ideal lines, and their r^2 values were high. However, the data points of the SDNN, LF, HF, and LF:HF ratio are dispersed, and their r^2 values are relatively low.

In addition, Bland–Altman analysis was performed to investigate the agreement between the HR and HRV parameters extracted from the ring and ECG. Figure 5 shows the Bland–Altman plots. The mean bias and 95% CI are shown in Figure 5 and Table 2. The ring (on average) overestimated pNN50, LF, and HF values but underestimated the other parameters. The HR, RMSSD, AVNN, and pNN50 had narrow 95% CIs; however, SDNN, LF, HF, and LF:HF ratio had relatively wider 95% CIs.

We also demonstrated the nocturnal HR and HRV parameters of one participant (randomly selected) in Figure 6. The parameters were obtained from the 5-minute segments. Figure 6 shows how the collected parameters from the Oura Ring (in red) and from the ECG (in green) vary throughout the night. As indicated, there are missing values, particularly in the frequency domain parameters, because of the removal of low-quality segments of the ECG or IBI signals.

Table 2. Pearson correlation coefficient, *P* values, 95% CI, and mean bias for heart rate (HR) and HR variability parameters between Ring and Shimmer3 in 5-minute window time.

Parameters	Pearson correlation coefficient	*P* value	95% CI	Mean bias
HR	0.99341	<.001	−2.81 to 1.93	−0.44
RMSSD[a]	0.91502	<.001	−44.07 to 14.13	−14.97 ms
SDNN[b]	0.51772	<.001	−88.45 to 86.52	−0.96 ms
AVNN[c]	0.82486	<.001	−210.01 to 183.24	−13.39 ms
pNN50[d]	0.76024	<.001	−0.23 to 0.35	0.06
LF[e] band	0.42401	<.001	−1758.9 to 1806.12	23.61 ms^2
HF[f] band	0.62734	<.001	−1423.92 to 1484.38	30.23 ms^2
LF:HF ratio	0.35455	<.001	−2.53 to 2.31	−0.11

[a]RMSSD: root mean square of successive differences between normal heartbeats.

[b]SDNN: SD of normal beat-to-beat intervals.

[c]AVNN: average of normal heartbeat intervals.

[d]pNN50: percentage of successive beat-to-beat intervals that differ by more than 50 ms.

[e]LF: low frequency.

[f]HF: high frequency.

Figure 4. The scatter plots and regression analysis of the nocturnal heart rate and heart rate variability parameters collected from the Oura Ring and Shimmer electrocardiography in 5-minute segments. The regression lines and ideal lines are indicated in red and black, respectively. AVNN: average of all normal-to-normal intervals; ECG: electrocardiography; HF: high frequency; HR: heart rate; LF: low frequency; pNN50: percentage of successive beat-to-beat intervals that differ by more than 50 ms; PPG: photoplethysmography; RMSSD: root mean square of successive differences; SDNN: SD of beat-to-beat intervals.

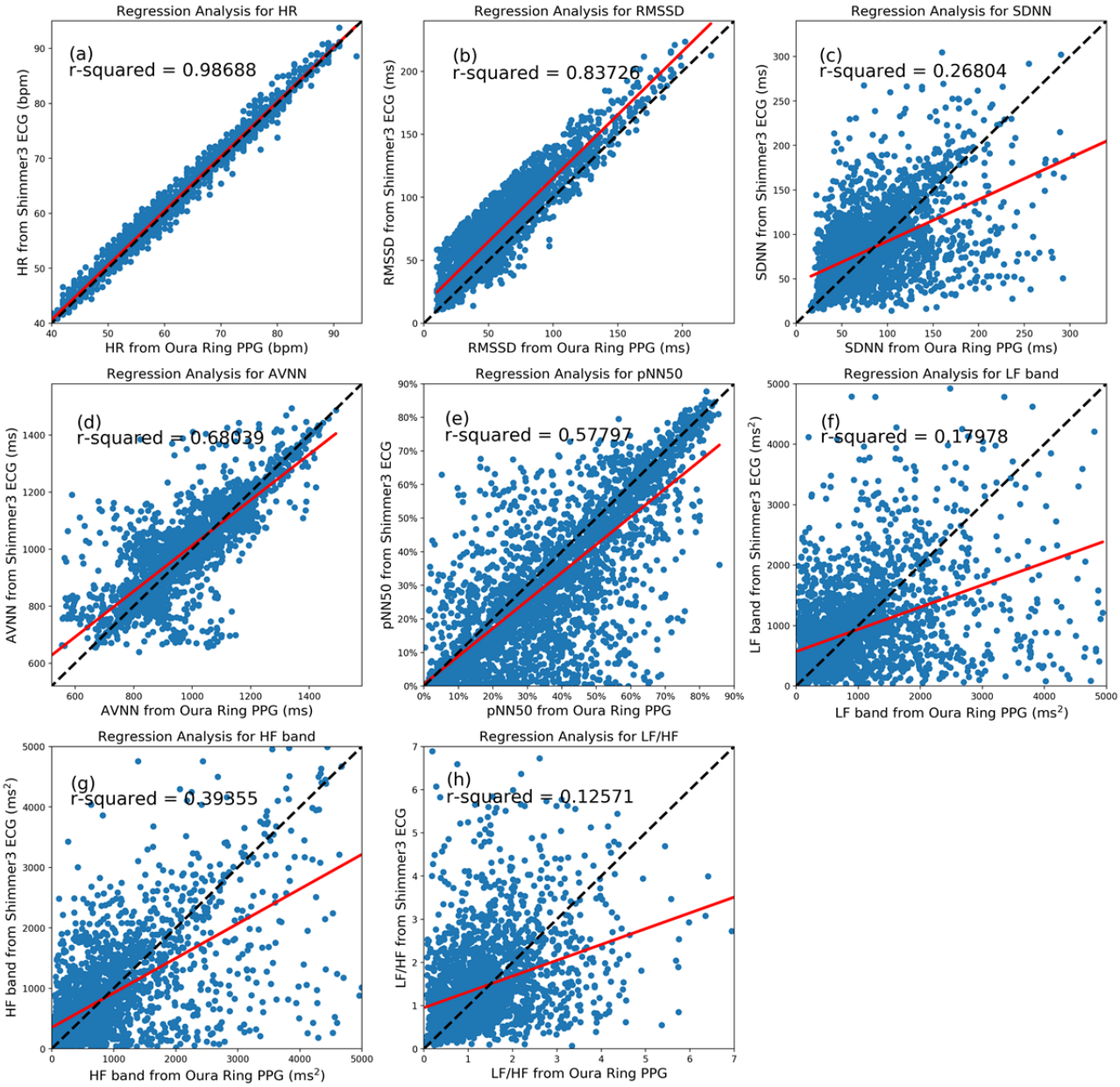


Figure 5. The Bland–Altman plots of the nocturnal heart rate and heart rate variability parameters in 5-minute segments obtained by the Oura Ring and Shimmer electrocardiography. AVNN: average of all normal-to-normal intervals; HF: high frequency; HR: heart rate; LF: low frequency; pNN50: percentage of successive beat-to-beat intervals that differ by more than 50 ms; PPG: photoplethysmography; RMSSD: root mean square of successive differences; SDNN: SD of beat-to-beat intervals.

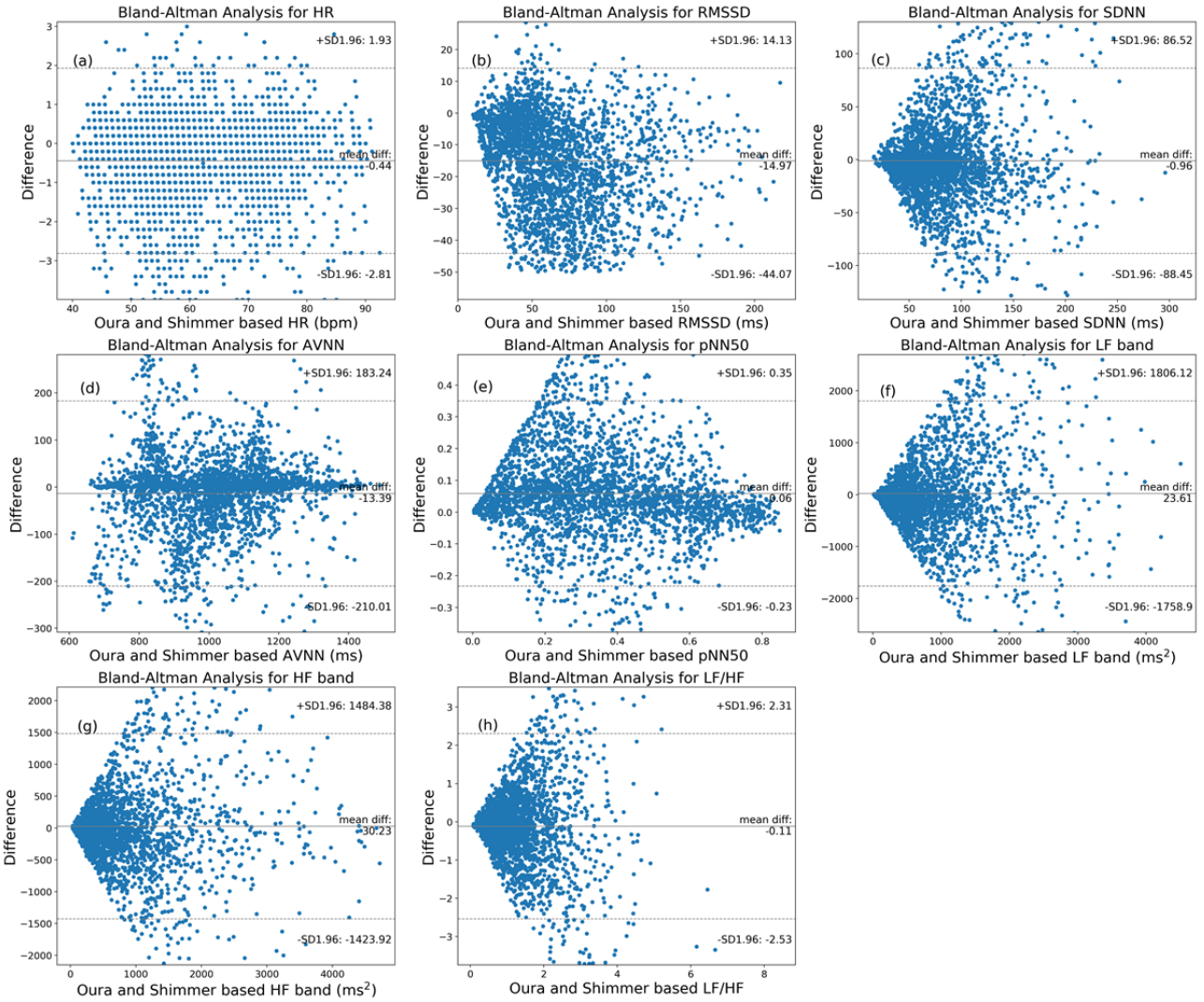
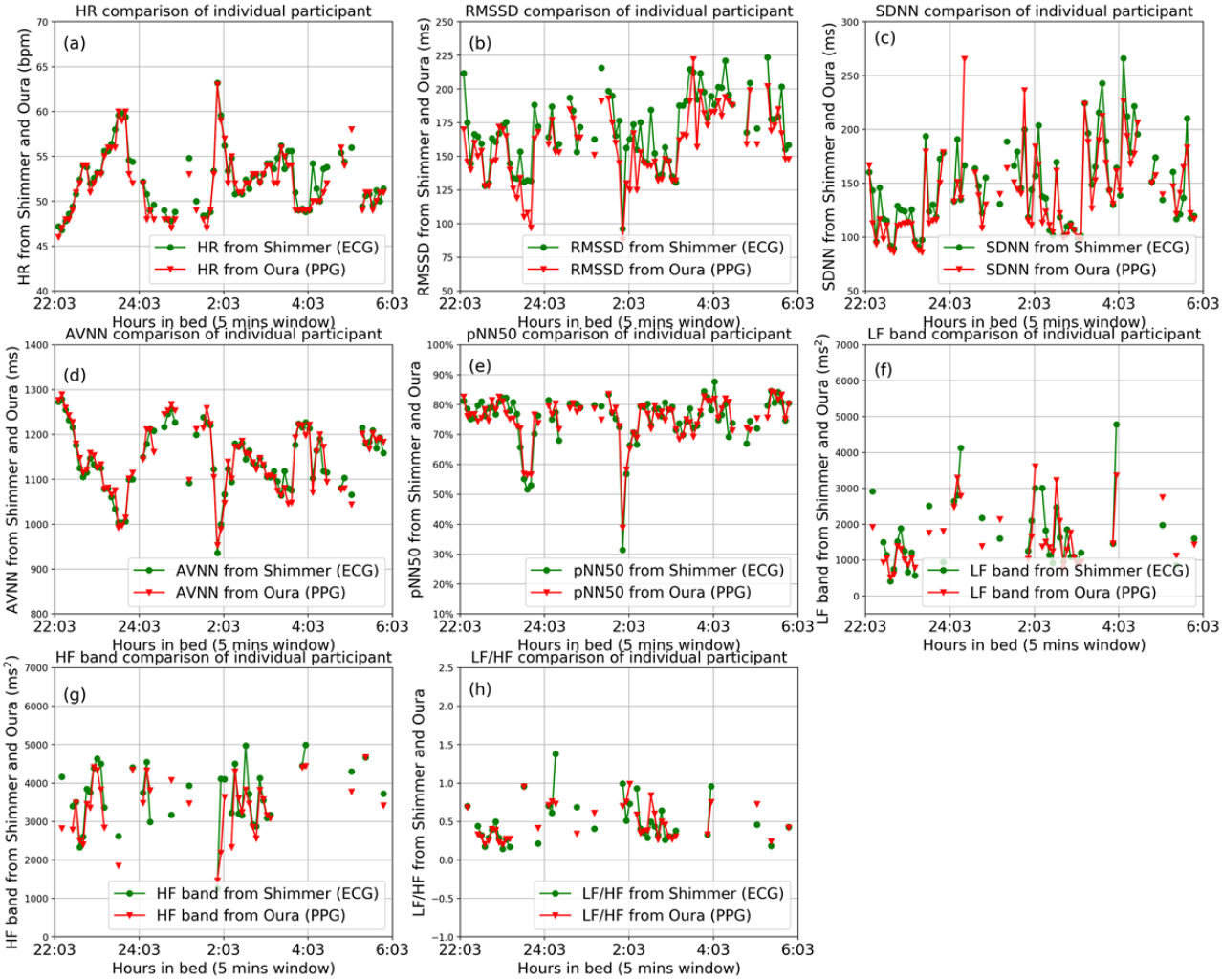


Figure 6. The nocturnal heart rate and heart rate variability parameters of one participant during a night sleep event. The data are extracted from the Oura Ring (in red) and Shimmer electrocardiography (in green). AVNN: average of all normal-to-normal intervals; ECG: electrocardiography; HF: high frequency; HR: heart rate; LF: low frequency; pNN50: percentage of successive beat-to-beat intervals that differ by more than 50 ms; PPG: photoplethysmography; RMSSD: root mean square of successive differences; SDNN: SD of beat-to-beat intervals.



Comparisons of Average HR and HRV Parameters of Ring and Shimmer3 During Sleep Time

We compared the average HR and HRV parameters per night sleep for the 2 devices to evaluate the overall errors. In this regard, we first extracted the correlation between the average HR and HRV parameters using the Pearson correlation test. Table 3 shows the Pearson correlation coefficients and corresponding *P* values. The correlation values of HR and RMSSD were very close to 1. Therefore, there were very strong positive correlations between the 2 devices. The AVNN, SDNN, pNN50, LF, and HF were higher than 0.8, showing high positive correlations between the 2 devices. There was also a moderately positive relationship in the LF:HF ratio.

We also used regression analysis to evaluate the average HR and HRV parameters throughout the night sleep period. Figure 7 illustrates the HR and HRV samples per night, the regression lines in red, the r^2 values, and the ideal lines (ie, y=x) in black. The r^2 values of the HR and RMSSD were greater than 0.9, indicating that the samples were near the regression lines. The r^2 values of SDNN, AVNN, pNN50, LF, and HF represent good fits. However, the r^2 value for the LF:HF ratio was 0.49.

In addition, we used the Bland–Altman analysis to investigate the differences between the average parameters per night sleep from these 2 devices (Figure 8). Table 3 shows the mean bias and 95% CI values. The results show that, on average, the ring overestimates pNN50, LF, and HF but underestimates the other parameters. Moreover, the 95% CIs of the HR, RMSSD, AVNN, and pNN50 were narrow, whereas the values were relatively wider for the SDNN and frequency domain parameters. These results are in accordance with the results presented in the previous section—Comparisons of HR and HRV Parameters of Ring and Shimmer3 in 5-Minute Time Windows.

Table 3. Pearson correlation coefficient, *P* values, 95% CI, and mean bias for the average heart rate (HR) and HR variability parameters per night collected from the Oura Ring and Shimmer3.

Parameter	Pearson correlation coefficient	*P* value	95% CI	Mean bias
HR	0.99968	<.001	−0.92 to 0.03	−0.44
RMSSD[a]	0.96210	<.001	−33.29 to 1.53	−15.88 ms
SDNN[b]	0.88469	<.001	−25.88 to 24.37	−0.76 ms
AVNN[c]	0.88010	<.001	−153.75 to 133.64	−10.05 ms
pNN50[d]	0.91251-	<.001	−0.1 to 0.22	0.06
LF[e] band	0.82916	<.001	−535.08 to 570.17	17.54 ms^2
HF[f] band	0.92585	<.001	−542.39 to 598.06	27.83 ms^2
LF:HF ratio	0.69837	<.001	−0.98 to 0.78	−0.1

[a]RMSSD: root mean square of successive differences between normal heartbeats.

[b]SDNN: SD of normal beat-to-beat intervals.

[c]AVNN: average of normal heartbeat intervals.

[d]pNN50: percentage of successive beat-to-beat intervals that differ by more than 50 ms.

[e]LF: low frequency.

[f]HF: high frequency.

Figure 7. The scatter plots and regression analysis of the average heart rate and heart rate variability parameters (which are collected from the Oura Ring and Shimmer electrocardiography) per night sleep time. The regression lines and ideal lines are indicated in red and black, respectively. AVNN: average of all normal-to-normal intervals; ECG: electrocardiography; HF: high frequency; HR: heart rate; LF: low frequency; pNN50: percentage of successive beat-to-beat intervals that differ by more than 50 ms; PPG: photoplethysmography; RMSSD: root mean square of successive differences; SDNN: SD of beat-to-beat intervals.

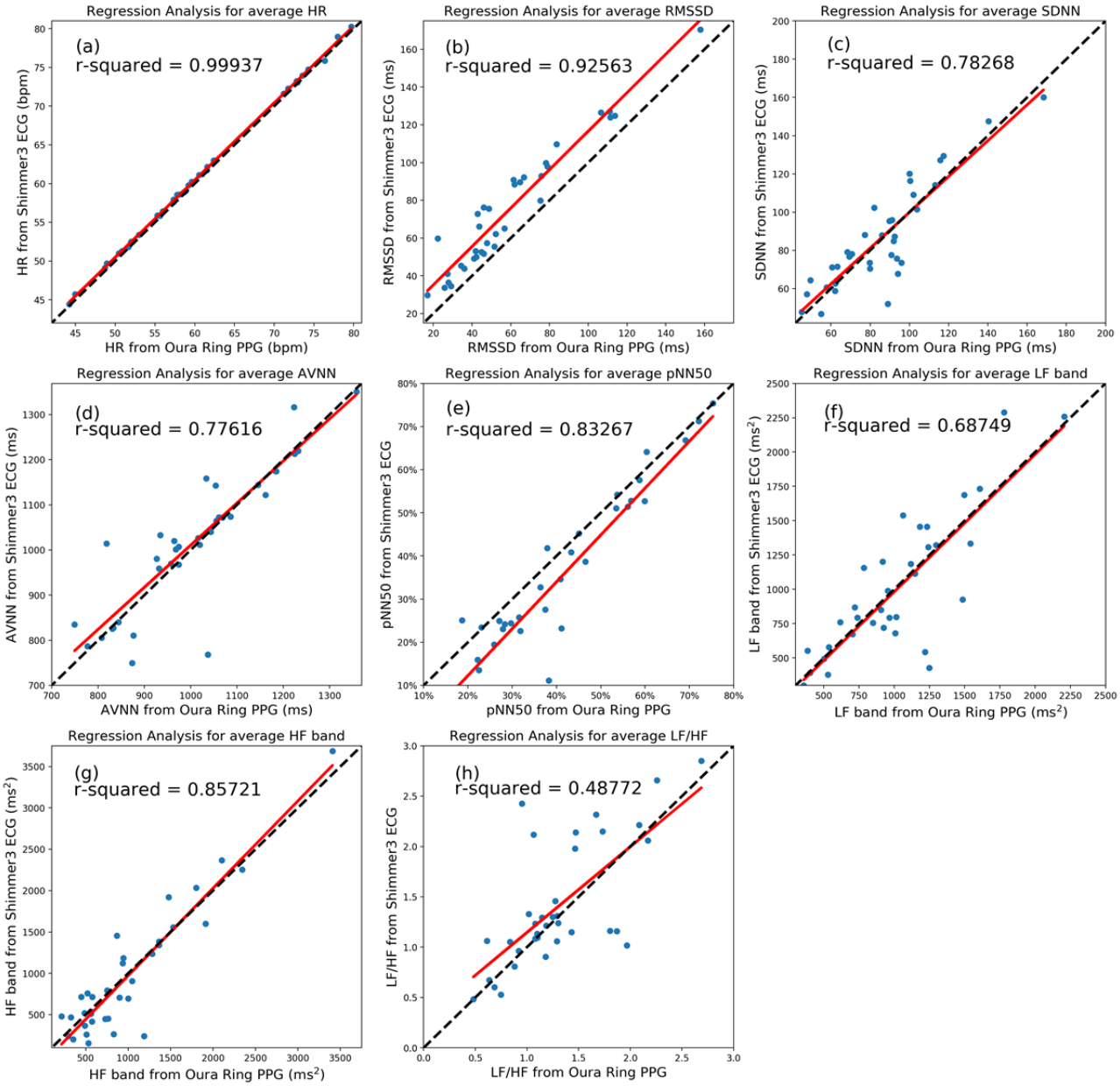
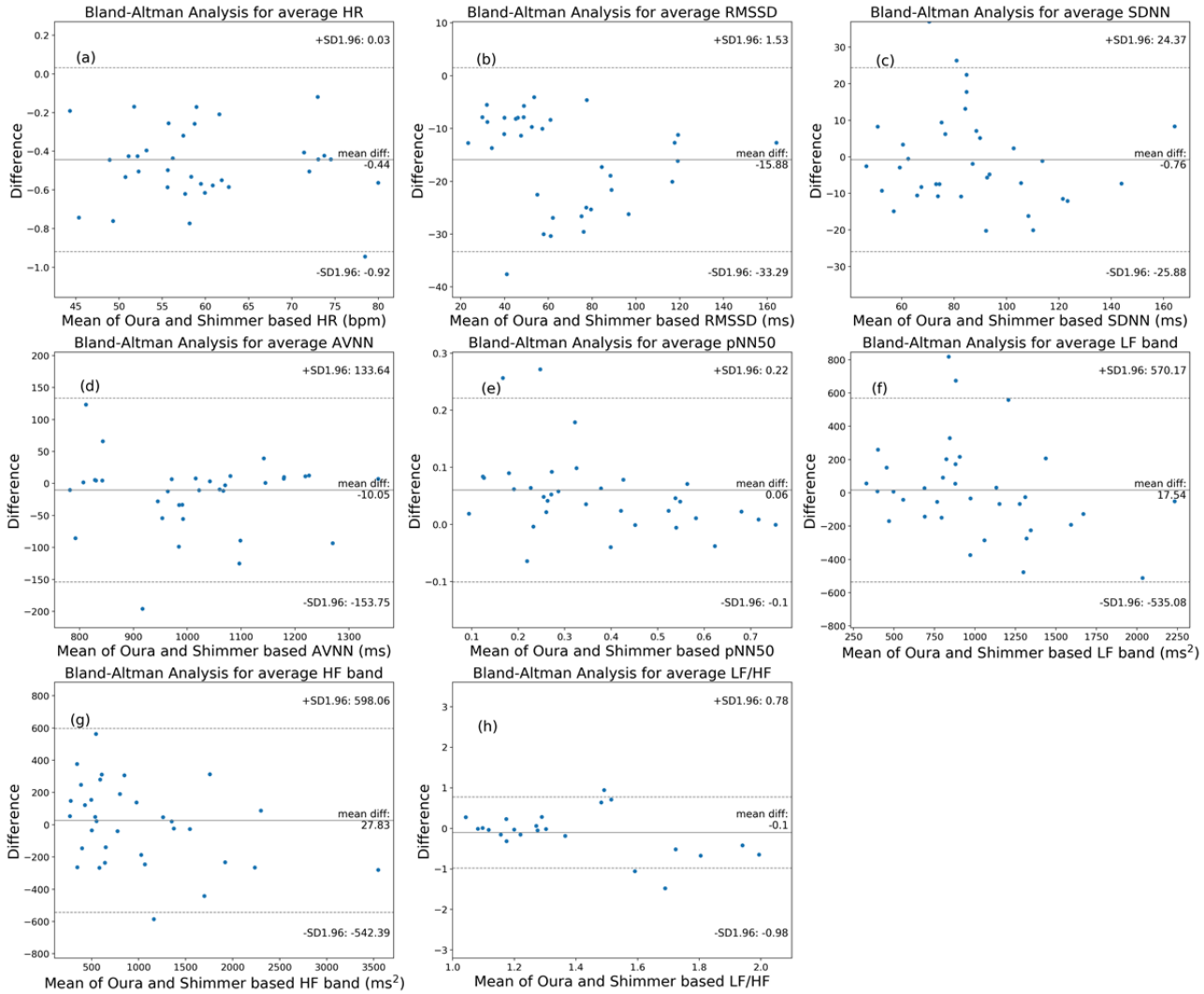


Figure 8. The Bland–Altman plots of the average heart rate and heart rate variability parameters (which are obtained by the Oura Ring and Shimmer electrocardiography) per night sleep time. AVNN: average of all normal-to-normal intervals; HF: high frequency; HR: heart rate; LF: low frequency; pNN50: percentage of successive beat-to-beat intervals that differ by more than 50 ms; PPG: photoplethysmography; RMSSD: root mean square of successive differences; SDNN: SD of beat-to-beat intervals.



Discussion

Principal Findings

In our analysis, we first validated the parameters extracted from the 5-minute PPG segments. 5-minute HRV recording, also known as short-term HRV analysis, is a measurement standard for extracting HRV parameters, such as RMSSD, SDNN, LF, and HF [4]. The LF:HF ratio is conventionally calculated via the 24-hour HRV recording [5]; however, it can also be collected in 5-minute recordings [4]. Our findings show relatively low mean biases for the HR and HRV parameters, where the Oura Ring overestimated pNN50, LF, and HF values but underestimated the other parameters. HR, RMSSD, AVNN, and pNN50 of the Oura Ring showed high positive correlations with the baseline, SDNN and HF showed moderate positive correlations, and LF and LF:HF ratio had low positive correlations.

However, the error variances of the parameters were different. The parameters provided by the Oura Ring dashboard (ie, HR and RMSSD) showed a relatively lower error variance compared with the HRV parameters extracted from the IBI signals. The error of HR is lower than that of RMSSD, which is in accordance with other studies showing that RMSSD is more sensitive to motion artifacts [35,49]. Among the parameters extracted from the IBI signals, AVNN and pNN50 showed moderate error rates compared with the baseline. However, SDNN, LF, HF, and LF:HF ratio had relatively higher error rates. The findings of the frequency domain parameters follow those of other studies, which show that these parameters are more sensitive to noise [30].

We also compared the average HR and HRV parameters during nighttime sleep. This comparison evaluates the long-term trends of HRV parameters and shows the validity of the parameters in a per-night analysis [34]. The mean biases per night were low, which is in accordance with the 5-minute recording analysis. In contrast, the error variances of the average values per night were considerably lower. This can be explained by the variance decrease because of averaging of the independent measurements. HR, RMSSD, AVNN, pNN50, SDNN, LF, and HF indicated high positive correlations, and the LF:HF ratio had a moderate positive correlation. To summarize, the average HR and HRV

parameters per night were relatively more accurate than the parameters extracted from 5-minute segments. Our results showed that the Oura Ring could accurately measure HR and RMSSD in both the 5-minute and average-per-night tests. The ring provided acceptable nocturnal AVNN, pNN50, HF, and SDNN accuracy in the average-per-night test but not in the 5-minute test. In contrast, the LF and LF:HF ratio of the ring had high error rates in both tests.

Comparison With Previous Studies

To the best of our knowledge, this is the first study to evaluate different HRV parameters of the Oura Ring in comparison with a standard ECG device. Kinnunen et al [34,35] focused on assessing the HR and RMSSD of the Oura Ring. In the 5-minute segment analysis, the HR and RMSSD were highly accurate. We obtained a higher r^2 for HR and a lower r^2 for RMSSD. Moreover, for the average-per-night analysis, we obtained almost the same r^2 for HR but lower r^2 for RMSSD. Our results indicate a narrower 95% CI and a smaller mean bias difference for average HR, and a wider 95% CI and a greater mean bias difference for average RMSSD.

Limitations

This study is limited to the nocturnal HR and HRV parameters, as the Oura Ring only provides the HR, RMSSD, and IBI values during sleep [34]. Future work should include the assessment of HR and HRV parameters during awake time. The PPG signals, and subsequently the parameters, might be distorted because of artifacts when the users engage in various activities and environments [50]. Such an evaluation is essential when using the ring in remote health monitoring and wellness tracking apps.

A total of 46 individuals participated in this home-based study, and data from 35 individuals were included in the analysis. However, this study was restricted to overnight data collection. Our future work will consider assessing the ring over the data collected over several days or weeks. This validation will provide a higher confidence level for the validity of the reported HR and HRV parameters.

Another limitation is the lack of generalizability of the results to nonhealthy individuals, as the study only included healthy participants. Recent studies have shown that the validity of wearable devices may be different for different population groups [51,52]. For example, atrial fibrillation affects the heart rhythms (irregular beats) of PPG [3]. Therefore, both time and frequency domain HRV parameters of individuals with atrial fibrillation are not the same as those of healthy people [36]. Consequently, the accuracy of the PPG-based atrial fibrillation methods should be investigated separately. Future directions for this study should include evaluating the PPG-based parameters acquired from individuals of different ages and with various health conditions.

Conclusions

In this study, we comprehensively evaluated the validity of the HR and HRV parameters collected by the Oura Ring. Our results showed low mean biases for the 8 parameters. In the 5-minute test, the error variances of the parameters were different. The parameters provided by the Oura Ring dashboard (ie, HR and RMSSD) showed relatively low error variance compared with the HRV parameters extracted from the IBI signals. HR, RMSSD, AVNN, and pNN50 of the ring indicated high positive correlations with the baseline values; SDNN and HF had moderate positive correlations; and LF and LF:HF ratio showed low positive correlations. In contrast, the average-per-night test indicated considerably lower error variances than the 5-minute test for all parameters. The Oura Ring was capable of accurately measuring HR and RMSSD in both the 5-minute and average-per-night tests. The ring indicated acceptable nocturnal AVNN, pNN50, HF, and SDNN accuracy in the average-per-night test but not in the 5-minute test. In contrast, the LF and LF:HF ratio of the ring had high error rates in both tests. Future work should include assessing the HR and HRV of the ring in long-term monitoring of population groups with different health conditions.

Acknowledgments

The authors would like to thank Elisa Lankinen, Mohsen Saei Dehghan, Bushra Zafar, and Henrika Merenlehto for contributing to the data collection. The authors acknowledge Oura Health Ltd for access to the data. This work was supported in part by the United States National Science Foundation through the WiFiUS grant CNS-1702950.

Conflicts of Interest

None declared.

References

1. McGaughey J, Alderdice F, Fowler R, Kapila A, Mayhew A, Moutray M. Outreach and Early Warning Systems (EWS) for the prevention of intensive care admission and death of critically ill adult patients on general hospital wards. Cochrane Database Syst Rev 2007 Jul 18(3):CD005529. [doi: 10.1002/14651858.CD005529.pub2] [Medline: 17636805]
2. Anzanpour A, Azimi I, Gotzinger M, Rahmani AM, TaheriNejad N, Liljeberg P, et al. Self-awareness in remote health monitoring systems using wearable electronics. In: Proceedings of the Design, Automation & Test in Europe Conference & Exhibition (DATE). 2017 Presented at: Design, Automation & Test in Europe Conference & Exhibition (DATE); Mar 27-31, 2017; Lausanne, Switzerland. [doi: 10.23919/DATE.2017.7927146]

3. McCraty R, Shaffer F. Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk. Glob Adv Health Med 2015 Jan;4(1):46-61 [FREE Full text] [doi: 10.7453/gahmj.2014.073] [Medline: 25694852]
4. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health 2017 Sep 28;5:258 [FREE Full text] [doi: 10.3389/fpubh.2017.00258] [Medline: 29034226]
5. Song T, Qu XF, Zhang YT, Cao W, Han BH, Li Y, et al. Usefulness of the heart-rate variability complex for predicting cardiac mortality after acute myocardial infarction. BMC Cardiovasc Disord 2014 May 01;14:59 [FREE Full text] [doi: 10.1186/1471-2261-14-59] [Medline: 24886422]
6. Guzzetti S, La Rovere MT, Pinna G, Maestri R, Borroni E, Porta A, et al. Different spectral components of 24 h heart rate variability are related to different modes of death in chronic heart failure. Eur Heart J 2005 Feb;26(4):357-362. [doi: 10.1093/eurheartj/ehi067] [Medline: 15618038]
7. Benichou T, Pereira B, Mermillod M, Tauveron I, Pfabigan D, Maqdasy S, et al. Heart rate variability in type 2 diabetes mellitus: a systematic review and meta-analysis. PLoS One 2018 Apr 2;13(4):e0195166 [FREE Full text] [doi: 10.1371/journal.pone.0195166] [Medline: 29608603]
8. Cardoso CR, Moraes RA, Leite NC, Salles GF. Relationships between reduced heart rate variability and pre-clinical cardiovascular disease in patients with type 2 diabetes. Diabetes Res Clin Pract 2014 Oct;106(1):110-117. [doi: 10.1016/j.diabres.2014.07.005] [Medline: 25145830]
9. Euliano TY, Michalopoulos K, Singh S, Gregg AR, Del Rio M, Vasilopoulos T, et al. Photoplethysmography and heart rate variability for the diagnosis of preeclampsia. Anesth Analg 2018 Mar;126(3):913-919 [FREE Full text] [doi: 10.1213/ANE.0000000000002532] [Medline: 28991110]
10. Stein PK, Pu Y. Heart rate variability, sleep and sleep disorders. Sleep Med Rev 2012 Feb;16(1):47-66. [doi: 10.1016/j.smrv.2011.02.005] [Medline: 21658979]
11. Burton AR, Rahman K, Kadota Y, Lloyd A, Vollmer-Conna U. Reduced heart rate variability predicts poor sleep quality in a case-control study of chronic fatigue syndrome. Exp Brain Res 2010 Jul;204(1):71-78. [doi: 10.1007/s00221-010-2296-1] [Medline: 20502886]
12. Hjortskov N, Rissén D, Blangsted AK, Fallentin N, Lundberg U, Søgaard K. The effect of mental stress on heart rate variability and blood pressure during computer work. Eur J Appl Physiol 2004 Jun;92(1-2):84-89. [doi: 10.1007/s00421-004-1055-z] [Medline: 14991326]
13. Han HJ, Labbaf S, Borelli JL, Dutt N, Rahmani AM. Objective stress monitoring based on wearable sensors in everyday settings. J Med Eng Technol 2020 May;44(4):177-189. [doi: 10.1080/03091902.2020.1759707] [Medline: 32589065]
14. Biomedical Signal Analysis, Second Edition. Hoboken, New Jersey: John Wiley & Sons; 2015.
15. What is HRV? OURA. URL: https://ouraring.com/blog/what-is-heart-rate-variability/ [accessed 2021-12-19]
16. Chatterjee A, Kumar Roy U. PPG based heart rate algorithm improvement with Butterworth IIR Filter and Savitzky-Golay FIR Filter. In: Proceedings of the 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech). 2018 Presented at: 2nd International Conference on Electronics, Materials Engineering & Nano-Technology (IEMENTech); May 4-5, 2018; Kolkata, India. [doi: 10.1109/iementech.2018.8465225]
17. Chatterjee A. Non-invasive CardioVascular monitoring. J Environ Sci Comput Sci Eng Technol 2018 Feb;7(1):33-47. [doi: 10.24214/jecet.b.7.1.03347]
18. Chatterjee A, Prinz A. Image analysis on fingertip video to obtain PPG. Biomed Pharmacol J 2018 Dec 21;11(4):1811-1827. [doi: 10.13005/bpj/1554]
19. Naeini EK, Azimi I, Rahmani AM, Liljeberg P, Dutt N. A real-time PPG quality assessment approach for healthcare internet-of-things. Procedia Comput Sci 2019;151:551-558. [doi: 10.1016/j.procs.2019.04.074]
20. Azimi I, Oti O, Labbaf S, Niela-Vilen H, Axelin A, Dutt N, et al. Personalized maternal sleep quality assessment: an objective IoT-based longitudinal study. IEEE Access 2019;7:93433-93447. [doi: 10.1109/access.2019.2927781]
21. Saarikko J, Niela-Vilen H, Ekholm E, Hamari L, Azimi I, Liljeberg P, et al. Continuous 7-month internet of things-based monitoring of health parameters of pregnant and postpartum women: prospective observational feasibility study. JMIR Form Res 2020 Jul 24;4(7):e12417 [FREE Full text] [doi: 10.2196/12417] [Medline: 32706696]
22. Laccetti AL, Slack Tidwell R, Sheth NP, Logothetis C, VanAlstine M. Remote patient monitoring using smart phone derived patient reported outcomes and Fitbit data to enable longitudinal predictive modeling in prostate cancer: feasibility results and lessons on platform development. J Clin Oncol 2019 May 20;37(15_suppl):e18068. [doi: 10.1200/jco.2019.37.15_suppl.e18068]
23. Introducing the new Oura Ring Generation 3. OURA. URL: https://ouraring.com/ [accessed 2021-12-19]
24. UCSF partners with Oura smart ring to study early detection of COVID-19. FIERCE Healthcare. URL: https://www.fiercehealthcare.com/tech/ucsf-launches-study-to-use-wearable-data-from-oura-ring-for-early-covid-19-detection [accessed 2021-12-19]
25. Oura partners with the NBA as the season resumes. OURA. URL: https://blog.ouraring.com/partners-with-the-nba/ [accessed 2021-12-19]
26. DeepBeat: a multi-task deep learning approach to assess signal quality and arrhythmia detection in wearable devices. ArXiv.org. 2020. URL: https://arxiv.org/abs/2001.00155 [accessed 2021-12-19]

27. von Rosenberg W, Chanwimalueang T, Adjei T, Jaffer U, Goverdovsky V, Mandic DP. Resolving Ambiguities in the LF/HF Ratio: LF-HF scatter plots for the categorization of mental and physical stress from HRV. Front Physiol 2017;8:360 [FREE Full text] [doi: 10.3389/fphys.2017.00360] [Medline: 28659811]
28. Devos D, Kroumova M, Bordet R, Vodougnon H, Guieu JD, Libersa C, et al. Heart rate variability and Parkinson's disease severity. J Neural Transm (Vienna) 2003 Sep;110(9):997-1011. [doi: 10.1007/s00702-003-0016-8] [Medline: 12928836]
29. Jarchi D, Salvi D, Velardo C, Mahdi A, Tarassenko L, Clifton DA. Estimation of HRV and SpO2 from wrist-worn commercial sensors for clinical settings. In: Proceedings of the IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN). 2018 Presented at: IEEE 15th International Conference on Wearable and Implantable Body Sensor Networks (BSN); Mar 4-7, 2018; Las Vegas, NV, USA. [doi: 10.1109/bsn.2018.8329679]
30. Hernando D, Roca S, Sancho J, Alesanco A, Bailón R. Validation of the Apple watch for heart rate variability measurements during relax and mental stress in healthy subjects. Sensors (Basel) 2018 Aug 10;18(8):2619 [FREE Full text] [doi: 10.3390/s18082619] [Medline: 30103376]
31. Morelli D, Bartoloni L, Colombo M, Plans D, Clifton DA. Profiling the propagation of error from PPG to HRV features in a wearable physiological-monitoring device. Healthc Technol Lett 2018 Apr;5(2):59-64 [FREE Full text] [doi: 10.1049/htl.2017.0039] [Medline: 29750114]
32. Hahnen C, Freeman CG, Haldar N, Hamati JN, Bard DM, Murali V, et al. Accuracy of vital signs measurements by a smartwatch and a portable health device: validation study. JMIR Mhealth Uhealth 2020 Feb 12;8(2):e16811 [FREE Full text] [doi: 10.2196/16811] [Medline: 32049066]
33. Asgari Mehrabadi M, Azimi I, Sarhaddi F, Axelin A, Niela-Vilén H, Myllyntausta S, et al. Sleep tracking of a commercially available smart ring and smartwatch against medical-grade actigraphy in everyday settings: instrument validation study. JMIR Mhealth Uhealth 2020 Nov 02;8(10):e20465 [FREE Full text] [doi: 10.2196/20465] [Medline: 33038869]
34. Kinnunen H, Koskimäki H. 0312 The HRV Of The Ring - Comparison of nocturnal HR and HRV between a commercially available wearable ring and ECG. Sleep 2018 Apr;41(suppl_1):A120 [FREE Full text] [doi: 10.1093/sleep/zsy061.311]
35. Kinnunen H, Rantanen A, Kenttä T, Koskimäki H. Feasible assessment of recovery and cardiovascular health: accuracy of nocturnal HR and HRV assessed via ring PPG in comparison to medical grade ECG. Physiol Meas 2020 May 07;41(4):04NT01. [doi: 10.1088/1361-6579/ab840a] [Medline: 32217820]
36. Chen J, Wasmund SL, Hamdan MH. Back to the future: the role of the autonomic nervous system in atrial fibrillation. Pacing Clin Electrophysiol 2006 Apr;29(4):413-421. [doi: 10.1111/j.1540-8159.2006.00362.x] [Medline: 16650271]
37. Shimmer3 ECG Unit. Shimmer. URL: https://www.shimmersensing.com/products/shimmer3-ecg-sensor#applications-tab [accessed 2021-12-19]
38. Burns A, Greene BR, McGrath MJ, O'Shea TJ, Kuris B, Ayer SM, et al. SHIMMER™ – A wireless sensor platform for noninvasive biomedical research. IEEE Sensors J 2010 Sep;10(9):1527-1534. [doi: 10.1109/JSEN.2010.2045498]
39. Petelczyc M, Gierałtowski JJ, Żogała-Siudem B, Siudem G. Impact of observational error on heart rate variability analysis. Heliyon 2020 May;6(5):e03984 [FREE Full text] [doi: 10.1016/j.heliyon.2020.e03984] [Medline: 32462091]
40. Lin CT, Chang KC, Lin CL, Chiang CC, Lu SW, Chang SS, et al. An intelligent telecardiology system using a wearable and wireless ECG to detect atrial fibrillation. IEEE Trans Inform Technol Biomed 2010 May;14(3):726-733. [doi: 10.1109/titb.2010.2047401]
41. Peltola MA. Role of editing of R-R intervals in the analysis of heart rate variability. Front Physiol 2012;3:148 [FREE Full text] [doi: 10.3389/fphys.2012.00148] [Medline: 22654764]
42. How Oura measures respiratory rate from your finger. OURA. URL: https://ouraring.com/respiratory-rate [accessed 2021-12-19]
43. Pan J, Tompkins WJ. A real-time QRS detection algorithm. IEEE Trans Biomed Eng 1985 Mar;32(3):230-236. [doi: 10.1109/TBME.1985.325532] [Medline: 3997178]
44. Hamilton P. Open source ECG analysis. In: Proceedings of the Computers in Cardiology. 2002 Presented at: Computers in Cardiology; Sep 22-25, 2002; Memphis, TN, USA. [doi: 10.1109/cic.2002.1166717]
45. Orphanidou C, Bonnici T, Charlton P, Clifton D, Vallance D, Tarassenko L. Signal-quality indices for the electrocardiogram and photoplethysmogram: derivation and applications to wireless monitoring. IEEE J Biomed Health Inform 2015 May;19(3):832-838. [doi: 10.1109/JBHI.2014.2338351] [Medline: 25069129]
46. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, SciPy 1.0 Contributors. Author Correction: SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):352 [FREE Full text] [doi: 10.1038/s41592-020-0772-5] [Medline: 32094914]
47. Machine Learning in Python. Scikit-Learn. URL: https://scikit-learn.org/stable/ [accessed 2021-12-19]
48. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. In: Proceedings of the Python in Science Conference. 2010 Presented at: Python in Science Conference; Jun 28 - Jul 3, 2010; Austin, Texas. [doi: 10.25080/majora-92bf1922-011]
49. Pietilä J, Tolonen J, Jimison HB, Pavel M, Mehrang S, Helander E, et al. Evaluation of the accuracy and reliability for photoplethysmography based heart rate and beat-to-beat detection during daily activities. In: EMBEC & NBC 2017. Singapore: Springer; 2017.

50. Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photopletysmography. In: Proceedings of the 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. 2014 Presented at: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society; Aug 26-30, 2014; Chicago, IL, USA. [doi: 10.1109/embc.2014.6944419]
51. Cook JD, Prairie ML, Plante DT. Ability of the multisensory jawbone UP3 to quantify and classify sleep in patients with suspected central disorders of hypersomnolence: a comparison against polysomnography and actigraphy. J Clin Sleep Med 2018 May 15;14(5):841-848 [FREE Full text] [doi: 10.5664/jcsm.7120] [Medline: 29734975]
52. Cook JD, Prairie ML, Plante DT. Utility of the Fitbit Flex to evaluate sleep in major depressive disorder: a comparison against polysomnography and wrist-worn actigraphy. J Affect Disord 2017 Aug 01;217:299-305 [FREE Full text] [doi: 10.1016/j.jad.2017.04.030] [Medline: 28448949]

Abbreviations

AVNN: average of all normal heart beat intervals
ECG: electrocardiography
HF: high frequency
HR: heart rate
HRV: heart rate variability
IBI: interbeat interval
LF: low frequency
pNN50: percentage of successive beat-to-beat intervals that differ by more than 50 ms
PPG: photoplethysmography
RMSSD: root mean square of successive differences
SDNN: SD of beat-to-beat intervals

Edited by R Kukafka; submitted 26.01.21; peer-reviewed by RS Mahmoud, A Kanaan, A Chatterjee; comments to author 16.04.21; revised version received 08.06.21; accepted 08.11.21; published 18.01.22.

Please cite as:
Cao R, Azimi I, Sarhaddi F, Niela-Vilen H, Axelin A, Liljeberg P, Rahmani AM
Accuracy Assessment of Oura Ring Nocturnal Heart Rate and Heart Rate Variability in Comparison With Electrocardiography in Time and Frequency Domains: Comprehensive Analysis
J Med Internet Res 2022;24(1):e27487
URL: https://www.jmir.org/2022/1/e27487
doi:10.2196/27487
PMID:35040799

Original Paper

Harnessing Artificial Intelligence for Health Message Generation: The Folic Acid Message Engine

Ralf Schmälzle[1], PhD; Shelby Wilcox[1], BA, MA

Department of Communication, Michigan State University, East Lansing, MI, United States

Corresponding Author:
Ralf Schmälzle, PhD
Department of Communication
Michigan State University
404 Wilson Rd
East Lansing, MI, 48824
United States
Phone: 1 (517) 353 ext 6629
Email: schmaelz@msu.edu

Abstract

Background: Communication campaigns using social media can raise public awareness; however, they are difficult to sustain. A barrier is the need to generate and constantly post novel but on-topic messages, which creates a resource-intensive bottleneck.

Objective: In this study, we aim to harness the latest advances in artificial intelligence (AI) to build a pilot system that can generate many candidate messages, which could be used for a campaign to suggest novel, on-topic candidate messages. The issue of folic acid, a B-vitamin that helps prevent major birth defects, serves as an example; however, the system can work with other issues that could benefit from higher levels of public awareness.

Methods: We used the *Generative Pretrained Transformer-2* architecture, a machine learning model trained on a large natural language corpus, and fine-tuned it using a data set of autodownloaded tweets about *#folicacid*. The fine-tuned model was then used as a *message engine*, that is, to create new messages about this topic. We conducted a web-based study to gauge how human raters evaluate AI-generated tweet messages compared with original, human-crafted messages.

Results: We found that the *Folic Acid Message Engine* can easily create several hundreds of new messages that appear natural to humans. Web-based raters evaluated the clarity and quality of a human-curated sample of AI-generated messages as on par with human-generated ones. Overall, these results showed that it is feasible to use such a message engine to suggest messages for web-based campaigns that focus on promoting awareness.

Conclusions: The *message engine* can serve as a starting point for more sophisticated AI-guided message creation systems for health communication. Beyond the practical potential of such systems for campaigns in the age of social media, they also hold great scientific potential for the quantitative analysis of message characteristics that promote successful communication. We discuss future developments and obvious ethical challenges that need to be addressed as AI technologies for health persuasion enter the stage.

(J Med Internet Res 2022;24(1):e28858) doi:10.2196/28858

KEYWORDS

human-centered AI; campaigns; health communication; NLP; health promotion

Introduction

Background

Crafting a successful health message involves a mix of art and science. On the one hand, decades of research in linguistics and communication science provides numerous insights into coherent sentence structure, effective value propositions, and other language-specific factors that promote attention, memory, and engagement [1,2]. On the other hand, translating these abstract factors into an appealing, concrete message that could be used in a campaign still requires a leap that must be fueled by human creativity and intuition [3].

Moreover, as larger and longer-term campaigns usually require a multitude of diverse messages, message creation represents a resource-intensive bottleneck. Although computers are often able to increase efficiency related to message development (ie,

information gathering, collaborative environments, and graphic designs), the task of message creation was traditionally beyond their scope. Until a few years ago, computers could analyze a sentence and flag errors; however, they were not able to synthesize a meaningful new sentence. However, the latest advances in machine learning (ML) have equipped computers with the ability to generate language for messages that appear natural and readily comprehensible to humans. This work is highly relevant to health communication in general and campaigns in particular as it could be connected to the task of campaign message generation. Specifically, there is the possibility that language generation methods might help in creating and optimizing messages; however, as there has been little contact between the fields of health communication and language generation, more work is needed to examine this possibility. In this paper, we ask, "How feasible is it to automatically generate on-topic messages that could potentially promote awareness about specific health issues?"

In the following section, we first review how the internet and social networking sites have become part and parcel of health communication. Next, we present the health issue of folate or folic acid (FA) as our test case and highlight the need for campaigns to promote FA awareness. We then introduce recent studies on natural language generation (NLG). This leads to a study in which we use a data set of FA-related social media messages to train a *message engine*, which then generates hundreds of new messages about this topic. We evaluate the clarity and quality of these artificial intelligence (AI)–generated messages compared with human-generated content via a web-based study.

The Potential of Social Media Communication Campaigns to Raise Awareness About Specific Health Topics

Social media has become a key component of communication campaigns [4]. This development has enabled new forms of health communication that are more direct and engaging for users. Social media–based messaging has also led to unprecedented opportunities for optimizing and effectively delivering information to the masses via computationally heavy approaches such as A/B-testing, recommender systems, and targeting receiver characteristics or social network positions [5-10]. Social media can diffuse messages widely across the globe and deeply into interpersonal networks [11,12].

The role of social media within the health communication landscape is still evolving; however, almost all health campaigns have embraced social media as cost-effective and highly scalable channels for raising and sustaining public attention [4,13]. Specific health issues that are affected by a chronic lack of awareness can benefit substantially from social media awareness campaigns. This is perhaps most prominently demonstrated by the success of the amyotrophic lateral sclerosis ice water bucket challenge, which brought substantial public awareness to the disease of amyotrophic lateral sclerosis and encouraged donations to research.

Raising awareness and providing basic information is a critical first step toward prevention, considering that all health communication theories posit that if people are unaware of a specific health risk, they will not take preventive action [14]. Of course, many complex health behaviors involve factors beyond awareness and education, such as shifting norms and attitudes or persuading target audiences to engage in specific behaviors [3,15]. However, for some selected health problems, awareness and knowledge deficits can be the primary campaign goals [16,17], and for many others, raising awareness or keeping the issue on the public agenda [18] is at least a secondary goal. Thus, although we are not claiming that raising awareness is a cure-all solution, we consider it a critical first step for any message generation system.

The Case of FA Awareness

Simply raising awareness and providing essential knowledge can go a long way for prenatal health. Many people who are pregnant are intrinsically motivated to adhere to health recommendations if they know them, as can be measured via self-report and behavioral indicators, such as smoking quitting attempts, reduction in drinking, and changes in exercise and nutrition behaviors [19-21]. This includes eating a folate-rich diet (to minimize the risk of neural tube defects [NTDs]) or avoiding rare meat (risk of toxoplasmosis) and certain cheeses (to reduce the risk of listeria infection). However, awareness about FA and knowledge about FA-rich diets among women of childbearing age remain too low [22-24]. This is problematic as most pregnancies are only noticed after NTDs occur, such that once people learn about effective prevention behaviors during, for example, a physician's visit, it may be too late [25,26]. Therefore, the issue of FA awareness will serve as a proof-of-concept example to demonstrate the potential of AI-generated messages that could potentially be used to raise awareness by providing a steady feed of on-topic but novel messages in long-term health campaigns.

Folate is a vitamin that is required for the body to build cells [27]. Many fruits, vegetables, and other natural foods contain folate, and the synthetic form, *FA*, is used as a dietary supplement or food additive. A folate or FA deficiency during early pregnancy can lead to severe embryonal NTDs [28]. Thus, the World Health Organization and the Centers for Disease Control and Prevention (CDC) recommend that all women of childbearing age consume 400 µg of folate per day [29,30].

Lack of awareness about FA represents a problem that is, at least to some degree, preventable via health communication and education [22,31-33]. As argued above, most people who are pregnant are motivated to achieve FA supply but will only be able to follow the guidelines if they are aware of them in the first place. Moreover, the recommended steps are relatively easy to follow for many people. However, that is not to say that by simply raising awareness, all positive downstream effects would follow. As with most health behaviors, they are embedded in a biopsychosocial context, requiring, for example, availability of food or FA supplementation, cultural factors, and so forth. However, the basic problem constellation of lack of awareness, paired with a relatively high spontaneous motivation and high self-efficacy and response efficacy, suggests that mass media health campaigns are a promising strategy. Indeed, several

previous studies support that FA-related campaigns can produce positive effects [22,31,33].

New Challenges for Social Media Communication Campaigns

The key benefit of mass media campaigns on social media is that they can quickly disseminate messages into the homes of millions. Moreover, social media has made it much easier to reach specific audience demographics and keep track of relevant outcomes, such as whether messages are seen, shared, or commented on [3,4].

However, although campaigns are a highly scalable tool, conducting a successful campaign is still far from trivial and requires substantial monetary investment and sustained effort over a longer period [3,15,34]. When it is properly conducted, mass communication is highly cost-effective compared with other approaches [35,36], and most campaigns do not achieve high levels of exposure over a sustained period [35]. For instance, most campaigns only achieve approximately 40% exposure in their target audience [37], which naturally reduces their success as communication effects logically require that messages are seen in the first place [38]. Moreover, in the days of print, radio, and television campaigning, many campaigns comprised only a limited number of messages that were switched at a relatively slow rate (eg, weekly or monthly), if at all. Although the more professional campaigns nowadays feature *feeds* with dozens of messages, if not more [4], maintaining such an effort is very costly and requires dedicated personnel, formative processes, and summative evaluation throughout [39,40]. In summary, campaign creation and maintenance is an effortful business.

However, even campaigns that are executed skillfully have difficulties in reaching their audience. The low 40% exposure rate mentioned above came from a study published in 2004; however, since then, the internet has further exacerbated the competition for attention [41-43]. Specifically, the very nature of today's attention economy on social media requires that health communicators update content frequently. Otherwise, algorithms will downvote the content and make it less likely to be seen by the target audience [43,44]. Similarly, on the side of the audience, switching behavior and searching for novel information are very widespread [45]. In summary, the logistic effort needed to create campaign messages and ensure their constant dissemination, as well as the algorithmic and user-sided information selection decisions, pose challenges for maximizing the potential of health communication campaigns.

Overall, this situation invites new approaches that could help health communicators and practitioners create a large number of awareness messages, which could then be automatically scheduled to ensure a constant and variable feed of appealing and timely messages. The following section introduces how recent developments in NLG, a subfield of machine learning or AI research, offer a potential solution to message development and dissemination limitations.

The Potential of Language Models to Generate Domain-Specific Health Messages

Advances in natural language processing have made it increasingly possible to generate coherent messages [46]. Although enthusiasm and skepticism about using computers for text generation have waxed and waned for decades [47], the advances in the past decade have been particularly impressive as the quality of computer-generated texts is now at a level that makes it often indiscriminable from human-written text [48,49].

A model that attracted substantial public attention is the *Generative Pretrained Transformer-2* (GPT-2) [50]. In brief, GPT-2 is a deep learning–based ML model that performs expertly across several language-related tasks, such as text translation and summarization, question answering, and text generation [51,52]. Approximately 40 GB of data from >8 million webpages were used to train the basic model. GPT-2 comes in 4 sizes ranging from 124M, 355M, 774M, and 1.5B parameters. Humans generally find the output of GPT-2's text generations authentic and interesting. Notably, the model is publicly available and can be adapted to many text-based tasks, such as summarization, question answering, or generation.

Pretrained language models can be fine-tuned to specific domains [53]. Fine-tuning is a form of transfer learning in which an ML model trained on domain-general data is retrained on further domain-specific data to adapt to its particularities. The possibility of using such fine-tuned language models to generate domain-specific text has already been demonstrated across different disciplines [54,55]. However, we are not aware of any effort to examine this in the context of health communication. Thus, the question is whether fine-tuning GPT-2 to the domain of FA messages will enable it to generate new messages that are of sufficient clarity and quality to be useful for a potential social media health campaign.

Present Study

This research examines the capability of language models to generate realistic messages about FA, which could serve as suggestions for a potential health campaign. Furthermore, we ask whether this is realistic in the context of public health communication, a situation often characterized by a lack of funds and computational resources. In brief, we use messages, or tweets, from the popular message-sharing platform Twitter to fine-tune a GPT-2 model. Although the same approach can be used with other social media platforms, Twitter offers a relatively straightforward, mainly text-based message format with a 280-character limit and easy access to existing messages, making it the most promising candidate for piloting such a system. After downloading messages and training the message engine, we use the fine-tuned model as a *message engine* to generate a large number of new FA-specific tweets. We then examine the characteristics of the generated messages to identify the preconditions for success and the current limitations of these generated messages. Finally, we wanted to know how AI-generated messages would be compared against human-generated messages. To this end, we conduct a web-based study in which human judges evaluate the AI-generated and human-generated tweets in terms of clarity and quality.

Methods

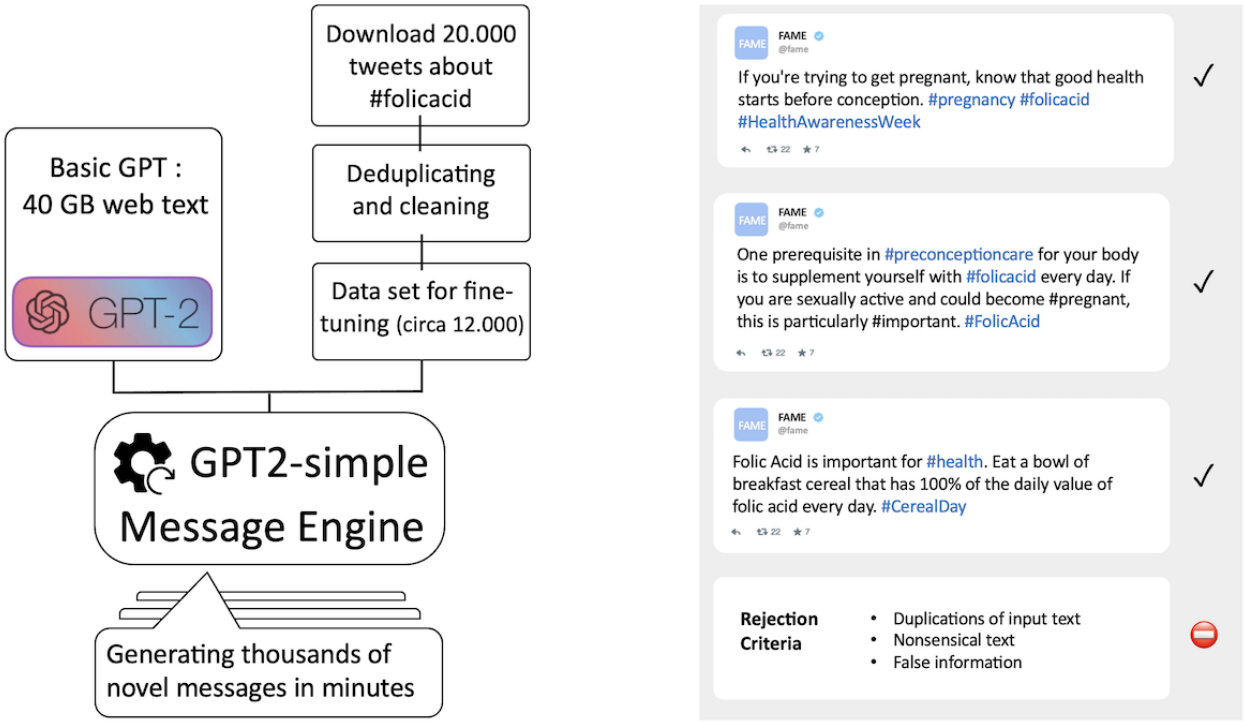
The FA Message Engine: Overall System Description

Overview

In this study, we harnessed the latest advances in AI to build a system that can generate a near-infinite number of health messages to promote FA awareness—the FA Message Engine. What will further be called the *message engine* is essentially an instantiation of the GPT-2 simple system, a Python package dedicated to fine-tuning OpenAI's GPT-2 text generation model [56]. We used the medium-sized GPT-2 model (355M hyperparameter versions trained on 40 GB of web text) and fine-tuned it with a data set of autodownloaded tweets about *#folicacid*. The resulting model was used to generate new messages (Figure 1).

Our specific steps are discussed in the following sections.

Figure 1. The left panel provides a schematic overview of the message engine construction and message generation workflow. The right panel illustrates a few examples of candidate messages. GPT: Generative Pretrained Transformer; GPT-2: Generative Pretrained Transformer-2.



Scraping Tweets for Model Retraining

To obtain a data set to fine-tune the model, we used the Twitter Intelligence Tool [57] to scrape a large number of tweets that mentioned *#folicacid* in their text. Specifically, we downloaded 25,304 tweets posted between 2010 and 2020 that mentioned *#folicacid* and extracted the raw text of the tweets. After removing duplicates, non-English tweets, and tweets that mainly promoted nutritional supplements (stopwords: buy, order, or sale), we ended up with a data set of 11,311 unique tweets for fine-tuning.

Fine Tuning the GPT-2

After downloading and cleaning the messages to create a data set for fine-tuning, the next step involved preparing and retraining the GPT-2 model. Specifically, we submitted the data set for fine-tuning to retrain the 355M GPT-2 model with recommended default settings of 2000 training steps and a learning rate of α=.0001. Fine-tuning was accomplished via Jupyter Notebook running Python 3 on a computer equipped with a graphics processing unit and executing the gpt2.finetune()-Method from the gpt-simple package [56]. On a standard graphics processing unit–equipped computer, fine-tuning a model of this size takes approximately 1 hour. We also conducted pilot experiments with other model sizes but chose to put only the medium-sized 355M GPT-2 model for a user test. Larger models require advanced hardware, whereas the medium model can work with most cloud-based computing services available to end users. Larger models are also not recommended for generating short text messages, such as tweets. After training, the fine-tuned model, which constitutes the *message engine*, was saved to the disk.

Generating Candidate Tweets via the Message Engine

We used the *message engine* with a default temperature setting of t=0.7 to generate 1000 new tweets. Temperature settings influence the randomness of the textual output, with a lower temperature being less random. As for model size, we conducted pilot tests with different temperature settings but noted that higher settings (t=1.0) produced very incoherent output, and low settings (t=0.3 and t=0.5) led to text that was very close to the training data. Given that our goal was to test the engine's output in humans in terms of clarity and quality, we deemed it worthwhile to conduct user testing for this setting, which is also the recommended default setting as per gpt2-simple's documentary [56].

Evaluating AI-Generated Messages: Web-Based Study

To evaluate the clarity and quality of tweets generated by the *message engine* against human-generated tweets, we performed a web-based study. The procedure was devised based on the emerging guidelines for evaluating NLG studies [58] and is described in the following sections.

Message Selection

From the 1000 AI-generated tweets, we drew a random sample of 60 tweets. Next, a human editor curated these tweet suggestions and compiled them into a set of 30 tweets for the web-based study. The human editor rejected AI-tweet suggestions if they contained duplications from the input data, false information according to CDC guidelines, or problematic advice (see the following section for details). A second human curator confirmed this selection without contradictions.

In parallel, we drew a random sample of 30 tweets from a pool of >10,000 real-life tweets. This strategy was chosen as it is not feasible to evaluate thousands of tweets and as it most likely mimics how practitioners would use such a system [49]. Thus, this procedure yielded 2 sets of 30 tweets each—30 AI-generated messages that came from a pool of 60 randomly drawn samples and 30 human-generated messages from Twitter.

Participants

We recruited 150 young adults from a web-based pool at a large Midwestern university to evaluate these messages in terms of clarity and quality. Study participants received course credit as reimbursement for completing the short survey, which lasted approximately 20 minutes and was approved by the local institutional review board. Of the 150 young adults, after excluding data from participants who did not finish the survey or responded unrealistically fast and clicked through the survey, we ended up with a data set of 129 (86%) respondents (mean age 20 years; range 18-28 years). The sample was predominantly female (96/129, 74.4%). Although this sample was not intended to be representative of the population, our participants clearly belonged to the audience of a potential FA awareness campaign. Moreover, given that the goal was to evaluate message clarity and quality rather than message effects on attitudes or behavior, this sample is sufficient for this purpose.

A power analysis suggested that a sample of approximately 100 raters was sufficient to detect a small-to-moderate effect in terms of the mean difference in evaluations of AI-generated and human-generated tweets ($1\text{-}\beta=0.9$; $\alpha=.05$; $dz=0.3$) [59]. Moreover, message evaluation studies suggest that evaluations of individual messages stabilize after averaging data from approximately 25 to 30 raters per message [60], which we surpassed with this sample size.

Procedure

The survey was administered via Qualtrics software (Qualtrics International), and participants were asked to evaluate all messages regarding clarity and quality. Participants were told that the study's goal was to examine human evaluations of Twitter messages about FA or folate, such as whether they considered the messages adequate to raise awareness or educate audiences about this health issue. The test messages were presented randomly, and participants were unaware of whether they came from the pool of AI-generated or human-generated messages. Each message was evaluated on 2 questions, 1 focusing on message clarity ("Please evaluate this message in terms of whether it is clear and easy to understand.") and 1 on message quality ("How much do you agree that the content and quality of this message is appropriate to increase public knowledge about folic acid?"). Answers were collected using a 5-point Likert-style response format (*very clear* and *very unclear* and *strongly agree* and *strongly disagree*). At the end of the survey, participants were debriefed about the study's purpose and provided a link to the CDC's website for the most up-to-date information on FA.

Evaluating AI-Generated Messages: Computational Analyses

In addition to inspecting the AI-generated messages and performing a web-based evaluation study, we conducted several computational analyses. Specifically, we computed n-grams and inspected their distribution between AI- and human-generated messages, including visualizations as word clouds. Next, we performed topic modeling analyses to gain additional insights into the semantic structure. Topic modeling is a prominent method for identifying health topics in social media [61] or subtopics within a given health domain [62-64]. Specifically, we used the topicmodels package [65] within the R statistical software to compute the latent Dirichlet allocation topic models [66]. Finally, we assessed the semantic similarity of individual messages via the sentence-transformers package [67]. To this end, we transformed each message into a sentence embedding and compared different messages via cosine-vector similarity.

Results

Overview

We found that the fine-tuned GPT-2 model can act as a *message engine* by creating grammatically correct, coherent, and novel messages centered on the topics of FA, healthy nutrition, and pregnancy. In the following sections, we will first describe the insights gained during the overall procedure and qualitative characteristics of the generated output, followed by the web-based evaluation study's quantitative results.

Feasibility of the System and Qualitative Description of the AI-Generated Messages

Our overall research question focused on whether it is possible to fine-tune a language model such as GPT-2 to a specific health domain to build a *message engine*. The answer is that it is possible. As can be seen from the sample output in Figure 1, the *FA Message Engine* was able to generate 1000 tweets within a matter of minutes, most of which resembled authentic web-based messages in style and content.

We next asked whether training such a system is realistic in the context of public health, where computational resources and specialized coding skills are scarce. The answer to this question is that it is feasible and surprisingly easy to implement. Although developing the scraping, cleaning, and training procedure took

some time, now that the system is set up, it can be replicated with little effort. For instance, if we wanted to replace the topic of *#folicacid* with any other health issue, this can be done in >1 hour. The system is also relatively accessible, even to novice users, as long as they are able to execute Python notebooks. Such skill requires only little training, and it would be possible to build a user interface for the system such that the user only enters the topic or search term (eg, *#folicacid*) and, after fine-tuning and generation, receives a sample of 60 message suggestions.

Most critically, we were interested in the characteristics of the generated messages, to which we turn next.

First, we note that the vast majority of the AI-generated tweets appeared natural and contained many elements of the original input tweets that were scraped from Twitter. For instance, the system uses hashtags that co-occurred with the search term *#folicacid*, such as *#pregnancy, #vitamin, #foodfortification #folicacidawarenessweek,* or *#eathealthy*. Second, as with hashtags, the system also tagged accounts that appeared in the input data, such as *@CDC* or *@NHS* (note that by eliminating these accounts from the input data, such information can be suppressed if not wanted).

Another observation is that most of the generated tweets were rather engaging, enthusiastic, or upbeat. This impression may again arise as the input tweets contained elements such as prompts with exclamation marks ("Eat healthy *now*!" and "*Go Folic! Visit [URL]!*") or encouragement, all of which could be interpreted as *cues-to-action* or attempts to raise *self-efficacy* according to the Health Belief Model [68]. This characteristic is likely as GPT-2 was trained on outgoing links with high so-called *karma scores* [50], thereby selectively emphasizing the language that web-based audiences found interesting and engaging.

Beyond resembling the linguistic style and platform-specific cues that are characteristic of today's Twitter environment (eg, upbeat language and hashtags), we observed that the AI-generated tweets reflected the input data's topic distribution. For instance, input tweets could be categorized into several topical clusters, such as nutritional needs during pregnancy, the link between FA and NTDs, political advocacy for mandatory food fortification, and so forth. Most AI-generated messages could also be categorized into coarse topic clusters. Additional results and visualizations of n-grams, word clouds, and results from topic modeling can be found in Multimedia Appendix 1 [65-67,69-73].

Although the overall system and procedure proved feasible, and the quality of many messages appeared comparable with human-generated messages, we made several observations that point to current limitations.

A simple observation is that the system sometimes parrots the training data; that is, it contains either duplicates of raw tweets or specific formulations that appeared in the data set used for fine-tuning (eg, "If you are trying to get pregnant..." and "Thinking of trying for a baby..."). This issue is well-known and follows logically from the fact that language models are essentially giant statistical association machines, which will learn the information contained in the input data. From an intellectual property perspective, this issue can raise questions about the copyright of the generated output. However, in practice, it is easy to sort out such parrot generations through human supervision, n-gram matching, or paraphrase detection algorithms.

A second limiting observation is that even when not directly parroting the training data, many tweets are still *close to* individual input messages, for example, by mixing formulations such as *trying to get pregnant* and *thinking of having a baby* with various combinations like *start taking #folicacid* or *know that good health starts before conception*. This points to mere reformulations and permutations that are not very creative. Again, this represents a direct consequence of the way natural language models work and is thus not necessarily a severe limitation. In fact, variations of a common message on a health topic may improve a campaign's reach by preventing the content from being downvoted by algorithms that select for novelty. Slight variations may also be beneficial in improving existing messages by making them briefer or more engaging, and it is well-known that repeated exposure to messages improves awareness and retention [38,74-76]. Rather than a limitation, this reformulation strategy of message generation can help to more optimally exhaust the space of possible effective messages. Nevertheless, it is clear that mere rewording represents only a minor achievement in message creation.

A third observation is that some generated messages contained false statements about which foods contain which amounts of FA or what medical defects might occur. Although such tweets are easy to spot in practice, this is an actual limitation. A total of 2 factors may underlie such behavior. First, if the input data contain false or problematic health claims, which are pervasive on social media, then the system will learn them. In this case, the system should not be blamed; however, the curation of the data set for fine-tuning should be optimized. However, more critically, the state of the art of current language models implies that they will simply generate tweets that sound linguistically coherent but may not make sense. We have discussed this issue in the *Discussion* section, where we suggest advancements to the system.

These results suggest that human curation and supervision of AI-generated tweets are necessary for practical use cases. Thus, a campaign manager or team would need to monitor the retraining process and eliminate problematic content, which is also what we opted for to select tweets for the web-based evaluation study.

Quantitative Comparison of AI- and Human-Generated Messages

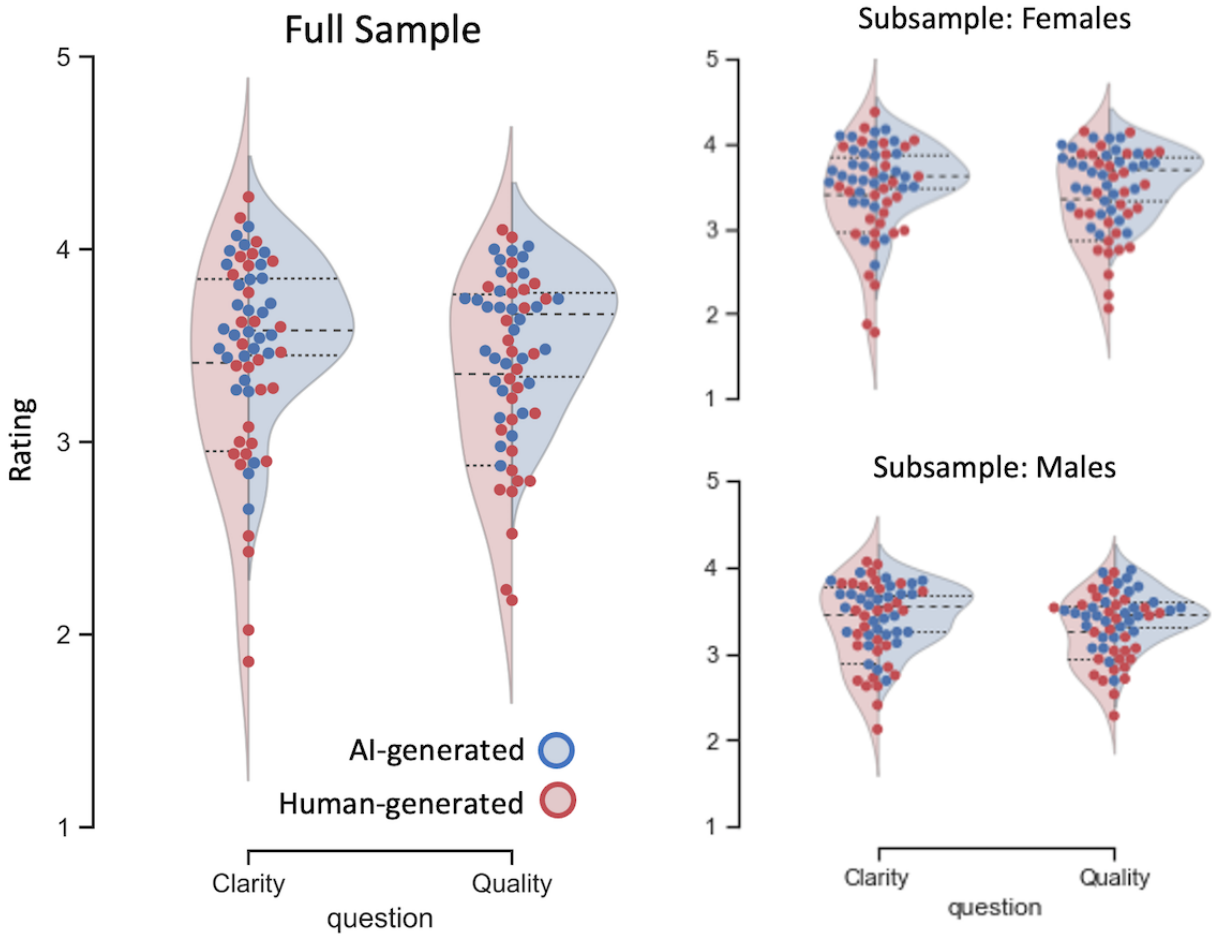
Table 1 shows the results of comparing AI-generated and human-generated tweets in terms of overall clarity or quality. Figure 2 illustrates the results graphically and provides further distributional information as well as analyses by subgroups. As can be seen, overall, the messages were rated as relatively clear and easy to understand, and participants found that their content and quality were appropriate for increasing public knowledge about FA (>3 on a 5-point scale).

Table 1. Means (SDs) from the web-based survey. Scores for message clarity and quality evaluations for 30 AI[a]- and 30 human-generated messages are shown, respectively.

Evaluations	AI-generated messages, mean (SD)	Human-generated messages, mean (SD)	*t* test (*df*)	P value
Clarity	3.58 (0.36)	3.34 (0.6)	1.97 (58)	.05
Quality	3.57 (0.32)	3.3 (0.53)	2.34 (58)	.02

[a]AI: artificial intelligence.

Figure 2. 30 artificial intelligence–generated tweets (blue) and 30 human-generated tweets (red) were evaluated in terms of perceived clarity and quality on a 5-point Likert-style scale. The results revealed very similar evaluations and minor average differences, with a considerable spread within each category. The right panels show analyses separated by gender. AI: artificial intelligence.



Statistical analysis revealed that the AI-generated and human-generated tweets were not rated as different in terms of how clear and easy they were to understand ($t_{clarity}$=1.97; P=.05). A small but statistically significant difference was found in the quality dimension ($t_{quality}$=2.34, P=.02). However, as can be seen from Figure 2, these mean differences were small in light of the variability, and thus, the effect was of a small size. We also zoomed in on the women participants' subgroup as the topic may be more relevant to them or become more relevant in the future with respect to pregnancy. As can be seen from Figure 2, the results were robust, and both subgroups (women and men) exhibited essentially the same pattern of results. However, we noted that, given the sample of college students, the topic of FA might not be very relevant to them, although there are also other health benefits of FA beyond the prevention of birth defects.

Next, we performed analyses at the level of the individual messages. As shown in Figure 2, the differences for both clarity and quality between individual messages were larger than the differences between categories (human- vs AI-generated). Indeed, performing item-wise analyses in which we compared ratings for single messages across raters using dependent-sample tests (because the same raters evaluated all messages) revealed that many messages were rated consistently higher than others. This pattern emerged both within human-generated and AI-generated messages, as well as across categories. Thus, many AI-generated messages were rated much higher than random human-generated messages.

Overall, we took these results as evidence that the *message engine* generated tweets that human raters evaluated as mainly equivalent to real Twitter messages.

Computational Analyses

In addition to the analyses of content (n-grams and topic modeling) and the human evaluation of clarity and quality, we wanted to examine the generated messages using computational methods. Specifically, we compared the 60 messages (30 AI-generated and 30 human-generated) using sentence Bidirectional Encoder Representations from Transformers (BERT), a modification of the pretrained bidirectional encoder representations from transformers model, to derive semantic sentence embeddings, which we then compared using cosine-similarity [67].

We found that across all messages, the average similarity was s=0.35. Within the 30 AI-generated messages, the average similarity was s_{AI}=0.37, and the average similarity between the 30 human-generated messages was s_{Human}=0.34. The average similarity between AI versus human messages was s_{AI} versus s_{Human}=0.35. Testing for differences between these computational indices of semantic similarities revealed no significant differences in any comparison (AI vs human and within- vs across-classes; all P>.08; for further details, see Multimedia Appendix 1 [65,66]). These results suggest that the sample of AI-generated messages is semantically similar to the sample of human-generated messages.

Discussion

Principal Findings

This study examined whether AI message generation technology can create candidate messages for use in social media health campaigns that focus primarily on raising awareness or increasing knowledge. We found that by retraining a GPT-2 model with thousands of tweets about FA, it is possible to build a *message engine* that can generate novel tweets, which could become part of an actual campaign. Human raters perceived these tweets as broadly similar in terms of clarity and quality to real-world messages. These results suggest that AI-assisted *message engines* could support campaign staff to create more efficient and possibly more effective campaigns for topics that are suitable for awareness-based messaging.

To our knowledge, this study is the first to demonstrate the potential of automated message generation in the context of health communication. Our results are generally positive, suggesting that the *FA Message Engine* can serve as a starting point for more sophisticated AI aides for message generation. Such systems can automatically offer thousands of messages that mimic the style and reflect the substance of existing health messages. Given that message creation is a resource-intensive bottleneck, we see significant application potential for such a system as a catalyst for human creativity [77].

Building a message engine for the topic of FA proved to be surprisingly easy. Our system made use of available tools [56] and could thus be transferred to contexts other than the issue of FA. Although this work did not intend to provide such a general purpose system for end users, it would be only a small step to deploy it as a web application as a turnkey solution.

Although the results of the web-based study demonstrate that the system output achieves good results, and the clarity ratings of AI-generated tweets are even significantly higher, we emphasize that our comparison strategy does not warrant the conclusion that AI-generated messages are superior to human-generated messages. Specifically, we compared a selection of 30 AI-generated messages against a sample of 30 real tweets, which were randomly drawn from a pool of >10,000 tweets. We opted for this procedure as our goal was to test the feasibility of AI-assisted message generation, which is the most realistic use case. In the following sections, we have discussed the significant limitations that currently prevent such a system from operating independently. However, the pool of human-curated AI-generated messages performed on par with or better than the standard tweets, and the analysis of semantic similarities did not reveal any difference. Thus, our approach suggests a simple strategy that might improve the quality of web-based content while saving the time and money of health communication practitioners.

Beyond the practical potential of such a *message engine* in the age of social media, the approach also offers considerable scientific potential. In particular, AI-based message engines might strengthen strategies to analyze message characteristics that underlie successful health communication [78-80]. In its current form, users of the message engine cannot influence the generated text's characteristics other than by what is fed in with the fine-tuning data set.

However, the natural language processing community [55,81,82] strives to gain more control over how the text is generated, and we see this as a promising next step. In particular, a limitation of the current system is that it does not incorporate any theory-based message design principles [1,83], such as barriers, cues to action, and norm or threat appeals. The fact that the current system learned to include some theory-compatible features, such as cues to action, shows promise in this regard; however, a more systematic approach is needed [84-86].

Ideally, this could then set off a virtuous cycle in which one could, via rapid iterations, gather feedback about specific message characteristics that are associated with targeted outcomes (eg, attention, awareness, and message sharing) and thus more clearly identify the message characteristics that facilitate individual outcomes [87]. As these characteristics become more accessible by linking objective message properties to large-scale outcomes, we might expect profound theoretical contributions from this otherwise applied system [88].

Along these lines, the most promising research direction is to fine-tune the fine-tuning process. We simply used the medium-sized GPT-2 model and fine-tuned the model with a set of tweets that were minimally screened. However, as with any manufacturing process, the quality of the input data determines the output. Thus, by fine-tuning the engine with often mediocre tweets, the AI-generated tweets were likely less potent than they would have been with a better training set.

In the future, we envision that one could curate a pool of high-quality tweets to serve as grade-A training material. An option, analogous to the strategy of training the GPT-2 base model with relatively more engaging text content, is to select

only those tweets about *#folicacid* that have been retweeted or liked. Another option is to bootstrap messages by having domain experts reword or craft theory-based examples. However, a challenge for this strategy is that fine-tuning requires large amounts of text—a few hundred examples are not enough. Overcoming this challenge is feasible with a large pool of quality input data. In addition, such a message pool could be used to train message engines for domains other than the narrow issue of FA.

We conclude this section by emphasizing again that the primary use case of such systems lies in boosting awareness for selected health problems where awareness is lacking or waning. At this point, a message engine system does not yet solve trickier health communication problems, such as the habitual nature of many negative health behaviors, addressing the socioecological embeddedness of such behaviors, or how to change health-related attitudes [89]. In principle, we see no reason why such systems could not be expanded to contexts beyond social media, especially as reliance on voice assistants such as Amazon's Alexa, Alibaba's AliGenie, or Apple's Siri for information increases. However, the *message engine* presented here is primarily intended for mass communication about public health issues that are affected by low awareness. For such health issues, we envision that this system can improve the cost/benefit ratio and overcome the *message-creation bottleneck* to avoid web-based content from being algorithmically downvoted as it is considered not fresh, dull, or unengaging.

Limitations, Risks, and Avenues for Future Research

This study demonstrates a positive application of NLG technology; however, some risks and limitations are worth mentioning.

A very basic limitation is that our focus was on demonstrating the feasibility of a message engine to generate messages that could potentially become part of a campaign; however, we did not actually conduct such a campaign. Thus, although we are confident that we showed that the generated messages—after going through the human content curation process—are on par with human-generated *baseline* messages, we did not actually show that these messages improved public health and especially not with regard to more distal outcomes such as attitude change and behavior. This should be the topic for future research.

Similarly, we note that our sample comprised college students who were not intended to be representative of the larger population. However, given that our focus was on evaluations of message clarity and quality rather than more idiosyncratically defined responses, this sample seems appropriate. This is also underscored by the fact that evaluations were highly consistent across subgroups of women and men raters. Nevertheless, future work on, for example, message effects on attitudes, beliefs, and other variables beyond basic clarity and quality should also focus on outcomes in specific health audiences, such as people who intend to have a child.

Regarding broader implications and risks, recent events in the political domain have highlighted the danger of algorithmic bots deployed to create or spread misinformation [90-93]. Several malicious actors seem to be using natural language generators to produce fake or divisive messages; thus, several empirical studies have examined the dangers of using NLG technology to create content that is harmful to society [49,94]. The same problems arise concerning the marketing of products that might harm health or use bots to promote certain brands [95-97].

Our study speaks to these issues by showing that it is also possible that benevolent actors can use NLG methods to promote public health. As with all technologies, risk and benefit are correlated, because otherwise the technology would be abandoned [98]. As such, we hope that our study will help explore the potential of AI as a force for promoting positive outcomes. However, this does not mean that we advocate for a laissez-faire strategy. Instead, a discussion of the ethical consequences of these technologies is needed and ongoing [99,100]. However, the field of health communication seems to be a particularly strong example of how human-centered AI could be used for social good.

Another risk and major limitation of the system is its lack of common sense knowledge. People who are not familiar with NLG technology are sometimes ambivalent about the idea of a *message engine*, finding it both magical and critical. Many also expect the systems to operate in a human-like fashion; however, this is not at all the case. On the surface, the generated tweets have a human-like look and feel to them; however, closer inspection reveals that GPT-2–style language models lack the deeper understanding and reasoning capacity that would be necessary for calling them intelligent.

Indeed, some AI-generated tweets are ludicrous, and others contain false information that is presented as fact [101]. For example, a tweet that emerged with little pretraining material was, "Make sure you drink 4 breads of #folicacid per day!" Such examples reflect a lack of common sense knowledge that one cannot drink bread. These examples arise only as GPT-2, although very sophisticated, ultimately boils down to a statistical association machine that links the domains nutrition and FA but does not have knowledge about fluid versus solid substances. The computer programs' inability to draw connections between categories or classify knowledge outside the training set has been a fundamental challenge since the early days of AI research [102-104].

This issue becomes particularly sensitive when generations contain wrong medical advice, such as "Take 4 lbs of FA per day!" Again, this reveals that the language model only picks up on statistical regularities in how words are used; however, it possesses no actual knowledge about pregnancy, nutrition, quantities, the developing fetus, and the causal relations between these concepts. These challenges are relatively easy to overcome with human supervision and insight. However, a trickier issue concerns issues in which the underlying knowledge is still evolving or uncertain. Such situations can provide fertile grounds for health myths or speculations about side effects. Such information will enter the message engine if it is present in the training data set, which again emphasizes the need for human curation.

In addition to the lack of domain knowledge about health and biology beyond that provided in the training set, we have already

pointed out that the system also has no theoretical understanding of communication and persuasion. The message engine only mimics and varies word use, albeit very eloquently. For a nonnative speaker who has to learn a foreign language for years, this skill may seem enviable; the ability to swiftly come up with 1000 sentences about FA may also impress and help health campaigners who spend hours coming up with 100 new candidate messages. However, the fact that such language models may talk without real understanding also means that they should only be used under the supervision of medical and communication experts. However, this is also true for more human-centric methods, such as focus groups or user-generated content.

Despite these limitations, the underlying technology can be expected to improve rapidly, and health communication researchers are well-advised to keep an eye on these developments. For instance, a successor model to GPT-2 has already been developed [105]. This model, called GPT-3, significantly improves some of the limitations that characterize GPT-2. Together with systems that are capable of generating persuasive arguments, selecting best-matching arguments for specific groups, and several other advances, we anticipate that the field of AI-assisted health message generation will see significant progress over the next decade [65,106-110].

Conclusions

To conclude, the *message engine* can generate candidate messages for human curators about selected health issues. This is relevant for issues where a lack of awareness is the primary problem, and a rich pool of social media messages is needed. At this stage, human supervision is necessary, and the technology, although very promising for content creation, requires control to select relevant content. Scientifically, this approach may promote a new wave of theoretical insights into the mechanisms of effective health messaging. We foresee that AI-generated messages for health promotion, education, and persuasion will become commonplace. It will be important for health communication researchers and practitioners to develop a strategy to use this technology positively.

Acknowledgments

The authors would like to thank Max Woolf for making the code for retraining the Generative Pretrained Transformer-2 available.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Inspection of word frequency, topic modeling results, and analysis of semantic similarity.
[DOCX File , 1928 KB - jmir_v24i1e28858_app1.docx]

References

1. Cho H. Health Communication Message Design: Theory and Practice. Thousand Oaks: SAGE; 2011.
2. Geeraerts D. Cognitive Linguistics: Basic Readings. Berlin: Walter de Gruyter; 2006.
3. Rice R, Atkin C. Public Communication Campaigns. Thousand Oaks: SAGE; 2012.
4. Shi J, Poorisat T, Salmon CT. The use of social networking sites (SNSs) in health communication campaigns: review and recommendations. Health Commun 2018 Jan 18;33(1):49-56. [doi: 10.1080/10410236.2016.1242035] [Medline: 27858464]
5. Kohavi R, Longbotham R. Online controlled experiments and A/B testing. In: Encyclopedia of machine learning and data mining. Boston: Springer; 2017.
6. Kim HS, Yang S, Kim M, Hemenway B, Ungar L, Cappella JN. An experimental study of recommendation algorithms for tailored health communication. Comput Commun Res 2019;1:103-129 [FREE Full text]
7. Matz SC, Kosinski M, Nave G, Stillwell DJ. Psychological targeting as an effective approach to digital mass persuasion. Proc Natl Acad Sci U S A 2017 Nov 28;114(48):12714-12719 [FREE Full text] [doi: 10.1073/pnas.1710966114] [Medline: 29133409]
8. Yom-Tov E, Shembekar J, Barclay S, Muennig P. Erratum: Author Correction: the effectiveness of public health advertisements to promote health: a randomized-controlled trial on 794,000 participants. NPJ Digit Med 2018 Aug 16;1:38 [FREE Full text] [doi: 10.1038/s41746-018-0047-z] [Medline: 31305589]
9. Gibson LA, Siegel L, Kranzler E, Volinsky A, O'Donnell MB, Williams S, et al. Combining crowd-sourcing and automated content methods to improve estimates of overall media coverage: theme mentions in e-cigarette and other tobacco coverage. J Health Commun 2019;24(12):889-899. [doi: 10.1080/10810730.2019.1682724] [Medline: 31718524]
10. Wang X, Chen L, Shi J, Peng T. What makes cancer information viral on social media? Comput Human Behav 2019 Apr;93:149-156. [doi: 10.1016/j.chb.2018.12.024]
11. Coronel JC, Ott JM, Hubner A, Sweitzer MD, Lerner S. How are competitive framing environments transformed by person-to-person communication? An integrated social transmission, content analysis, and eye movement monitoring approach. Commun Res 2020 Mar 01:009365022090359. [doi: 10.1177/0093650220903596]
12. Katz E, Lazarsfeld P, Roper E. Personal Influence: The Part Played By People in the Flow of Mass Communications. London: Routledge; 2017.

13. Zhao X. Health communication campaigns: a brief introduction and call for dialogue. Int J Nurs Sci 2020 Sep 10;7(Suppl 1):S11-S15. [doi: 10.1016/j.ijnss.2020.04.009] [Medline: 32995373]
14. Schmälzle R, Renner B, Schupp HT. Health risk perception and risk communication. Policy Insights Behav Brain Sci 2017 Aug 24;4(2):163-169. [doi: 10.1177/2372732217720223]
15. Pfau M, Parrott R. Persuasive Communication Campaigns. New York: Allyn and Bacon; 1992.
16. Dutka S, Colley R. DAGMAR, Defining Advertising Goals for Measured Advertising Results. Lincolnwood, IL: NTC Business Books; 1995.
17. Bettinghaus EP. Health promotion and the knowledge-attitude-behavior continuum. Preventive Med 1986 Sep;15(5):475-491. [doi: 10.1016/0091-7435(86)90025-3]
18. Jones K. Agenda setting in health and risk messaging. In: Oxford Research Encyclopedia of Communication. Oxford, United Kingdom: Oxford University Press; 2017.
19. Crozier S, Robinson S, Borland S, Godfrey K, Cooper C, Inskip H, et al. Do women change their health behaviours in pregnancy? Findings from the Southampton Women's Survey. Paediatr Perinat Epidemiol 2009 Sep;23(5):446-453 [FREE Full text] [doi: 10.1111/j.1365-3016.2009.01036.x] [Medline: 19689495]
20. Gaston A, Prapavessis H. Maternal-fetal disease information as a source of exercise motivation during pregnancy. Health Psychol 2009 Nov;28(6):726-733. [doi: 10.1037/a0016702] [Medline: 19916641]
21. Floyd RL, Rimer BK, Giovino GA, Mullen PD, Sullivan SE. A review of smoking in pregnancy: effects on pregnancy outcomes and cessation efforts. Annu Rev Public Health 1993;14:379-411. [doi: 10.1146/annurev.pu.14.050193.002115] [Medline: 8323595]
22. Rofail D, Colligs A, Abetz L, Lindemann M, Maguire L. Factors contributing to the success of folic acid public health campaigns. J Public Health (Oxf) 2012 Mar 03;34(1):90-99 [FREE Full text] [doi: 10.1093/pubmed/fdr048] [Medline: 21727078]
23. Medawar G, Wehbe T, Jaoude E. Awareness and use of folic acid among women of childbearing age. Ann Glob Health 2019 Apr 09;85(1):54 [FREE Full text] [doi: 10.5334/aogh.2396] [Medline: 30977622]
24. Green-Raleigh K, Carter H, Mulinare J, Prue C, Petrini J. Trends in folic acid awareness and behavior in the United States: the Gallup Organization for the March of Dimes Foundation surveys, 1995-2005. Matern Child Health J 2006 Sep;10(5 Suppl):S177-S182 [FREE Full text] [doi: 10.1007/s10995-006-0104-0] [Medline: 16823638]
25. Greene ND, Copp AJ. Neural tube defects. Annu Rev Neurosci 2014;37:221-242 [FREE Full text] [doi: 10.1146/annurev-neuro-062012-170354] [Medline: 25032496]
26. Cameron M, Moran P. Prenatal screening and diagnosis of neural tube defects. Prenat Diagn 2009 Apr;29(4):402-411. [doi: 10.1002/pd.2250] [Medline: 19301349]
27. Crider KS, Bailey LB, Berry RJ. Folic acid food fortification-its history, effect, concerns, and future directions. Nutrients 2011 Mar;3(3):370-384 [FREE Full text] [doi: 10.3390/nu3030370] [Medline: 22254102]
28. Geisel J. Folic acid and neural tube defects in pregnancy: a review. J Perinat Neonatal Nurs 2003;17(4):268-279. [doi: 10.1097/00005237-200310000-00005] [Medline: 14655787]
29. Folic Acid. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/ncbddd/folicacid/index.html [accessed 2022-01-10]
30. Periconceptional folic acid supplementation to prevent neural tube defects. World Health Organization. URL: https://www.who.int/elena/titles/folate_periconceptional/en/ [accessed 2022-01-10]
31. Chivu CM, Tulchinsky TH, Soares-Weiser K, Braunstein R, Brezis M. A systematic review of interventions to increase awareness, knowledge, and folic acid consumption before and during pregnancy. Am J Health Promot 2008;22(4):237-245. [doi: 10.4278/06051566R2.1] [Medline: 18421888]
32. Amitai Y, Fisher N, Haringman M, Meiraz H, Baram N, Leventhal A. Increased awareness, knowledge and utilization of preconceptional folic acid in Israel following a national campaign. Prev Med 2004 Oct;39(4):731-737. [doi: 10.1016/j.ypmed.2004.02.042] [Medline: 15351539]
33. van der Pal-de Bruin KM, de Walle HE, Jeeninga W, de Rover C, Cornel MC, de Jong-van den Berg LT, et al. The Dutch 'Folic Acid Campaign'--have the goals been achieved? Paediatr Perinat Epidemiol 2000 Apr;14(2):111-117. [doi: 10.1046/j.1365-3016.2000.00251.x] [Medline: 10791653]
34. Snyder LB. Health communication campaigns and their impact on behavior. J Nutr Educ Behav 2007;39(2 Suppl):S32-S40. [doi: 10.1016/j.jneb.2006.09.004] [Medline: 17336803]
35. Noar SM. An audience-channel-message-evaluation (ACME) framework for health communication campaigns. Health Promot Pract 2012 Jul;13(4):481-488. [doi: 10.1177/1524839910386901] [Medline: 21441207]
36. Hutchinson P, Wheeler J. The cost-effectiveness of health communication programs: what do we know? J Health Commun 2006;11 Suppl 2:7-45. [doi: 10.1080/10810730600973862] [Medline: 17148098]
37. Snyder LB, Hamilton MA, Mitchell EW, Kiwanuka-Tondo J, Fleming-Milici F, Proctor D. A meta-analysis of the effect of mediated health communication campaigns on behavior change in the United States. J Health Commun 2004;9 Suppl 1:71-96. [doi: 10.1080/10810730490271548] [Medline: 14960405]
38. Hornik RC. Exposure: theory and evidence about all the ways it matters. Soc Marketing Q 2016 Aug 01;8(3):31-37. [doi: 10.1080/15245000214135]

39. Farrelly M, Niederdeppe J, Yarsevich J. Youth tobacco prevention mass media campaigns: past, present, and future directions. Tob Control 2003 Jun;12 Suppl 1:i35-i47 [FREE Full text] [doi: 10.1136/tc.12.suppl_1.i35] [Medline: 12773784]
40. Roditis ML, Jones C, Dineva AP, Alexander TN. Lessons on addiction messages from "The real cost" campaign. Am J Prev Med 2019 Feb;56(2 Suppl 1):S24-S30 [FREE Full text] [doi: 10.1016/j.amepre.2018.07.043] [Medline: 30661522]
41. Simon H. Designing organizations for an information-rich world. In: Computers, Communications, and the Public Interest. Baltimore, Maryland, United States: Johns Hopkins Press; 1971.
42. Atchley P, Lane S. Chapter Four-cognition in the attention economy. In: Psychology of Learning and Motivation. New York: Academic Press; 2014.
43. Citton Y. The Ecology of Attention. London: John Wiley & Sons; 2017.
44. Attention Economy: Understanding the New Currency of Business. Brighton, Massachusetts: Harvard Business Review Press; 2002.
45. Jiang J, He D, Allan J. Searching, browsing, clicking in a search session: changes in user behavior by task over time. In: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval. 2014 Presented at: SIGIR '14: Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval; Jul 6-11, 2014; Gold Coast Queensland Australia. [doi: 10.1145/2600428.2609633]
46. Hirschberg J, Manning CD. Advances in natural language processing. Science 2015 Jul 17;349(6245):261-266. [doi: 10.1126/science.aaa8685] [Medline: 26185244]
47. Weizenbaum J. Computer Power and Human Reason: From Judgement to Calculation. New York, United States: W.H.Freeman & Co Ltd; 1976.
48. Seabrook J. Can a machine learn to write for The New Yorker? The New Yorker. URL: https://www.newyorker.com/magazine/2019/10/14/can-a-machine-learn-to-write-for-the-new-yorker [accessed 2022-01-10]
49. Kreps SE, McCain M, Brundage M. All the news that's fit to fabricate: AI-generated text as a tool of media misinformation. J Exp Polit Sci 2020 Nov 20:1-14. [doi: 10.1017/XPS.2020.37]
50. Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I. Language models are unsupervised multitask learners. OpenAI blog. URL: http://www.persagen.com/files/misc/radford2019language.pdf [accessed 2022-01-10]
51. Alammar J. The illustrated transformer. GitHub. URL: https://jalammar.github.io/illustrated-transformer/ [accessed 2022-01-10]
52. Radford A, Wu J, Amodei D, Amodei D, Clark J, Brundage M. Better language models and their implications. OpenAI Blog. URL: https://openai.com/blog/better-language-models/ [accessed 2022-01-10]
53. Peters M, Ruder S, Smith N. To tune or not to tune? Adapting pretrained representations to diverse tasks. arXiv. Preprint posted online March 14, 2019 [FREE Full text]
54. Frazier S, Al Nahian S, Riedl M, Harrison B. Learning norms from stories: a prior for value aligned agents. In: Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. 2019 Presented at: AIES '20: AAAI/ACM Conference on AI, Ethics, and Society; Feb 7-9, 2020; New York, USA URL: https://dl.acm.org/doi/proceedings/10.1145/3375627 [doi: 10.1145/3375627.3375825]
55. Ziegler D, Stiennon N, Wu J, Brown T, Radford A, Amodei D, et al. Fine-tuning language models from human preferences. arXiv. Preprint posted online September 18, 2019 [FREE Full text]
56. Woolf M. gpt-2-simple. Github Repository. URL: https://github.com/minimaxir/gpt-2-simple [accessed 2020-10-01]
57. Twint. GitHub. URL: https://github.com/twintproject/twint [accessed 2022-01-10]
58. van Gatt DL, van Wubben ME, Krahmer E. Best practices for the human evaluation of automatically generated text. In: Proceedings of the 12th International Conference on Natural Language Generation. 2019 Presented at: Proceedings of the 12th International Conference on Natural Language Generation; 2019; Tokyo, Japan. [doi: 10.18653/v1/w19-8643]
59. Erdfelder E, Faul F, Buchner A. GPOWER: a general power analysis program. Behav Res Methods Instrum Comput 1996 Mar;28(1):1-11. [doi: 10.3758/bf03203630]
60. Kim M, Cappella JN. An efficient message evaluation protocol: two empirical analyses on positional effects and optimal sample size. J Health Commun 2019;24(10):761-769 [FREE Full text] [doi: 10.1080/10810730.2019.1668090] [Medline: 31543057]
61. Paul MJ, Dredze M. Discovering health topics in social media using topic models. PLoS One 2014;9(8):e103408 [FREE Full text] [doi: 10.1371/journal.pone.0103408] [Medline: 25084530]
62. Pruss D, Fujinuma Y, Daughton AR, Paul MJ, Arnot B, Albers Szafir D, et al. Zika discourse in the Americas: a multilingual topic analysis of Twitter. PLoS One 2019 May 23;14(5):e0216922 [FREE Full text] [doi: 10.1371/journal.pone.0216922] [Medline: 31120935]
63. Zhou S, Zhao Y, Bian J, Haynos AF, Zhang R. Exploring eating disorder topics on Twitter: machine learning approach. JMIR Med Inform 2020 Oct 30;8(10):e18273 [FREE Full text] [doi: 10.2196/18273] [Medline: 33124997]
64. Boon-Itt S, Skunkan Y. Public perception of the COVID-19 pandemic on twitter: sentiment analysis and topic modeling study. JMIR Public Health Surveill 2020 Nov 11;6(4):e21978 [FREE Full text] [doi: 10.2196/21978] [Medline: 33108310]
65. Grün B, Hornik K. topicmodels: an R package for fitting topic models. J Stat Softw 2011;40(13):1-30 [FREE Full text] [doi: 10.18637/jss.v040.i13]
66. Blei D, Ng A, Jordan M. Latent dirichlet allocation. J Mach Learn Res 2003;3:993-1002 [FREE Full text]

67. Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks. ArXiv.org. 2019. URL: http://arxiv.org/abs/1908.10084 [accessed 2022-01-10]
68. Green E, Murphy E, Gryboski K. The health belief model. In: The Wiley Encyclopedia of Health Psychology. London: Wiley; 2020.
69. ldatuning: tuning of the latent dirichllocation models parameters. CRAN. URL: https://rdrr.io/cran/ldatuning/ [accessed 2022-01-10]
70. Cao J, Xia T, Li J, Zhang Y, Tang S. A density-based method for adaptive LDA model selection. Neurocomputing 2009 Mar;72(7-9):1775-1781. [doi: 10.1016/j.neucom.2008.06.011]
71. Arun R, Suresh V, Veni M, Narasimha MM. On finding the natural number of topics with Latent Dirichlet allocation: some observations. In: Advances in Knowledge Discovery and Data Mining. Berlin Heidelberg: Springer; 2010.
72. Deveaud R, SanJuan E, Bellot P. Accurate and effective latent concept modeling for ad hoc information retrieval. Document numérique 2014 Apr 30;17(1):61-84. [doi: 10.3166/dn.17.1.61-84]
73. Carlini N, Tramer F, Wallace E, Jagielski M, Herbert-Voss A, Lee K, et al. Extracting training data from large language models. arXiv. Preprint posted online December 14, 2020 [FREE Full text]
74. Baddeley A. Human Memory: Theory and Practice. New York: Psychology Press; 1997.
75. Neubaum G, Krämer NC, Alt K. Psychological effects of repeated exposure to elevating entertainment: an experiment over the period of 6 weeks. Psychol Popular Media 2020 Apr;9(2):194-207. [doi: 10.1037/ppm0000235]
76. Appleton K, Hemingway A, Rajska J, Hartwell H. Repeated exposure and conditioning strategies for increasing vegetable liking and intake: systematic review and meta-analyses of the published literature. Am J Clin Nutr 2018 Oct 01;108(4):842-856. [doi: 10.1093/ajcn/nqy143] [Medline: 30321277]
77. Maiden N. Digital creativity support: designing AI to augment human creativity. In: The Human Position in an Artificial World: Creativity, Ethics and AI in Knowledge Organization. Würzburg, Germany: Ergon; 2019.
78. Cappella J. Perceived message effectiveness meets the requirements of a reliable, valid, and efficient measure of persuasiveness. J Commun 2018 Oct;68(5):994-997 [FREE Full text] [doi: 10.1093/joc/jqy044] [Medline: 30479403]
79. O'Keefe DJ. Message generalizations that support evidence-based persuasive message design: specifying the evidentiary requirements. Health Commun 2015;30(2):106-113. [doi: 10.1080/10410236.2014.974123] [Medline: 25470435]
80. Kim HS. How message features and social endorsements affect the longevity of news sharing. Digit J (Abingdon) 2021;9(8):1162-1183 [FREE Full text] [doi: 10.1080/21670811.2020.1811742] [Medline: 34900400]
81. Keskar N, McCann B, Varshney L, Xiong C, Socher R. CTRL: a conditional transformer language model for controllable generation. arXiv. Preprint posted online September 11, 2019 [FREE Full text]
82. Lin Z, Riedl M. Plug-and-blend: A framework for controllable story generation with blended control codes. arXiv. Preprint posted online March 23, 2021 [FREE Full text]
83. Harrington N. Persuasive health message design. In: Oxford Research Encyclopedia of Communication. Oxford: Oxford University Press; 2016.
84. Scott AJ. Persuasive Advertising: Evidence-based Principles. London: Palgrave Macmillan; 2010.
85. Harrington NG. Introduction to the special issue: message design in health communication research. Health Commun 2015;30(2):103-105 [FREE Full text] [doi: 10.1080/10410236.2014.974133] [Medline: 25470434]
86. Tan C, Niculae V, Danescu-Niculescu-Mizil C, Lee L. Winning arguments: interaction dynamics and persuasion strategies in good-faith online discussions. arXiv. Preprint posted online February 2, 2016 [FREE Full text]
87. Boster FJ, Liu RW, Cheng Y, Kim W, Shaikh SJ. The Suasory force of sticky messages: an application to the application of sunscreen. Commun Stud 2017 Dec 26;69(1):4-22. [doi: 10.1080/10510974.2017.1414067]
88. Fishbein M, Cappella JN. The role of theory in developing effective health communications. J Commun 2006 Aug;56(Suppl 1):1-17. [doi: 10.1111/j.1460-2466.2006.00280.x]
89. Thompson TL, Schulz PJ. Health Communication Theory. London: John Wiley & Sons; 2021.
90. Woolley SC, Howard PN. Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media. Oxford: Oxford University Press; 2018.
91. Zellers R, Holtzman A, Rashkin H, Bisk Y, Farhadi A, Roesner F. Defending against neural fake news. arXiv. Preprint posted online May 29, 2019 [FREE Full text]
92. Subrahmanian V, Azaria A, Durst S, Kagan V, Galstyan A, Lerman K, et al. The DARPA Twitter bot challenge. Computer 2016 Jun;49(6):38-46. [doi: 10.1109/MC.2016.183]
93. Coronel J, Bucy E. A cognitive neuroscience perspective on political knowledge, misinformation, and memory for facts. In: The Handbook of Communication Science and Biology. London: Routledge; 2020.
94. Leib M, Köbis N, Rilke R, Hagens M, Irlenbusch B. The corruptive force of AI-generated advice. arXiv. Preprint posted online February 15, 2021 [FREE Full text]
95. Allem J, Ferrara E. Could social bots pose a threat to public health? Am J Public Health 2018 Aug;108(8):1005-1006. [doi: 10.2105/AJPH.2018.304512] [Medline: 29995482]
96. Shi W, Liu D, Yang J, Zhang J, Wen S, Su J. Social bots' sentiment engagement in health emergencies: a topic-based analysis of the COVID-19 pandemic discussions on Twitter. Int J Environ Res Public Health 2020 Nov 23;17(22):8701 [FREE Full text] [doi: 10.3390/ijerph17228701] [Medline: 33238567]

97. Unger JB, Rogers C, Barrington-Trimis J, Majmundar A, Sussman S, Allem J, et al. "I'm using cigarettes to quit JUUL": an analysis of Twitter posts about JUUL cessation. Addict Behav Rep 2020 Dec;12:100286 [FREE Full text] [doi: 10.1016/j.abrep.2020.100286] [Medline: 32637562]
98. Fischhoff B, Slovic P, Lichtenstein S, Read S, Combs B. How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits. Policy Sci 1978 Apr;9(2):127-152. [doi: 10.1007/BF00143739]
99. Solaiman I, Brundage M, Clark J, Askell A, Herbert-Voss A, Wu J. Release strategies and the social impacts of language models. arXiv. Preprint posted online August 24, 2019 [FREE Full text]
100. Gibney E. The battle for ethical AI at the world's biggest machine-learning conference. Nature 2020 Jan;577(7792):609. [doi: 10.1038/d41586-020-00160-y] [Medline: 31992885]
101. Frankfurt H. On Bullshit. Princeton: Princeton University Press; 2009.
102. Levesque H. Common sense, the Turing test,the quest for real AI. Boston: MIT Press; 2017.
103. Marcus G, Davis E. Rebooting AI: Building Artificial Intelligence We Can Trust. New York: Knopf Doubleday Publishing Group; 2019.
104. McCarthy J. Programs with common sense. Nature. URL: https://www.cs.rit.edu/~rlaz/files/mccarthy1959.pdf [accessed 2022-01-10]
105. Brown TB, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, et al. Language models are few-shot learners. arXiv. Preprint posted online May 28, 2020 [FREE Full text]
106. Hoyt RE, Snider D, Thompson C, Mantravadi S. IBM Watson analytics: automating visualization, descriptive, and predictive statistics. JMIR Public Health Surveill 2016 Oct 11;2(2):e157 [FREE Full text] [doi: 10.2196/publichealth.5810] [Medline: 27729304]
107. Ji H, Ke P, Huang S, Wei F, Zhu X, Huang M. Language generation with multi-hop reasoning on commonsense knowledge graph. arXiv. Preprint posted online September 24, 2020 [FREE Full text]
108. Bosselut A, Rashkin H, Sap M, Malaviya S, Celikyilmaz A, Choi Y. Comet: commonsense transformers for knowledge graph construction. arXiv. Preprint posted online June 12, 2019 [FREE Full text]
109. Mazuz K, Yom-Tov E. Analyzing trends of loneliness through large-scale analysis of social media postings: observational study. JMIR Ment Health 2020 Apr 20;7(4):e17188 [FREE Full text] [doi: 10.2196/17188] [Medline: 32310141]
110. Orji R, Moffatt K. Persuasive technology for health and wellness: state-of-the-art and emerging trends. Health Informatics J 2018 Mar;24(1):66-91 [FREE Full text] [doi: 10.1177/1460458216650979] [Medline: 27245673]

Abbreviations

AI: artificial intelligence
CDC: Centers for Disease Control and Prevention
FA: folic acid
GPT-2: Generative Pretrained Transformer-2
ML: machine learning
NLG: natural language generation
NTD: neural tube defect

Edited by R Kukafka; submitted 21.03.21; peer-reviewed by Z Su, R Rice, CY Lin, X Yang; comments to author 26.04.21; revised version received 21.05.21; accepted 21.11.21; published 18.01.22.

Please cite as:
Schmälzle R, Wilcox S
Harnessing Artificial Intelligence for Health Message Generation: The Folic Acid Message Engine
J Med Internet Res 2022;24(1):e28858
URL: https://www.jmir.org/2022/1/e28858
doi:10.2196/28858
PMID:35040800

Original Paper

Exploring the COVID-19 Pandemic as a Catalyst for Behavior Change Among Patient Health Record App Users in Taiwan: Development and Usability Study

Chinyang Henry Tseng[1], PhD; Ray-Jade Chen[2,3], MSc, MD; Shang-Yu Tsai[4], MSc; Tsung-Ren Wu[1], MSc; Woei-Jiunn Tsaur[1], PhD; Hung-Wen Chiu[4], PhD; Cheng-Yi Yang[5], PhD; Yu-Sheng Lo[4], PhD

[1]Department of Computer Science and Information Engineering, National Taipei University, New Taipei, Taiwan

[2]Department of Surgery, School of Medicine, College of Medicine, Taipei Medical University, Taipei, Taiwan

[3]Taipei Medical University Hospital, Taipei, Taiwan

[4]Graduate Institute of Biomedical Informatics, College of Medical Science and Technology, Taipei Medical University, Taipei, Taiwan

[5]Industrial Technology Research Institute, Hsinchu, Taiwan

Corresponding Author:
Yu-Sheng Lo, PhD
Graduate Institute of Biomedical Informatics
College of Medical Science and Technology
Taipei Medical University
250 Wuxing Street
Taipei, 110
Taiwan
Phone: 886 +886227361661
Email: Loyusen@tmu.edu.tw

Abstract

Background: During the COVID-19 pandemic, personal health records (PHRs) have enabled patients to monitor and manage their medical data without visiting hospitals and, consequently, minimize their infection risk. Taiwan's National Health Insurance Administration (NHIA) launched the My Health Bank (MHB) service, a national PHR system through which insured individuals to access their cross-hospital medical data. Furthermore, in 2019, the NHIA released the MHB software development kit (SDK), which enables development of mobile apps with which insured individuals can retrieve their MHB data. However, the NHIA MHB service has its limitations, and the participation rate among insured individuals is low.

Objective: We aimed to integrate the MHB SDK with our developed blockchain-enabled PHR mobile app, which enables patients to access, store, and manage their cross-hospital PHR data. We also collected and analyzed the app's log data to examine patients' MHB use during the COVID-19 pandemic.

Methods: We integrated our existing blockchain-enabled mobile app with the MHB SDK to enable NHIA MHB data retrieval. The app utilizes blockchain technology to encrypt the downloaded NHIA MHB data. Existing and new indexes can be synchronized between the app and blockchain nodes, and high security can be achieved for PHR management. Finally, we analyzed the app's access logs to compare patients' activities during high and low COVID-19 infection periods.

Results: We successfully integrated the MHB SDK into our mobile app, thereby enabling patients to retrieve their cross-hospital medical data, particularly those related to COVID-19 rapid and polymerase chain reaction testing and vaccination information and progress. We retrospectively collected the app's log data for the period of July 2019 to June 2021. From January 2020, the preliminary results revealed a steady increase in the number of people who applied to create a blockchain account for access to their medical data and the number of app subscribers among patients who visited the outpatient department (OPD) and emergency department (ED). Notably, for patients who visited the OPD and ED, the peak proportions with respect to the use of the app for OPD and ED notes and laboratory test results also increased year by year. The highest proportions were 52.40% for ED notes in June 2021, 88.10% for ED laboratory test reports in May 2021, 34.61% for OPD notes in June 2021, and 41.87% for OPD laboratory test reports in June 2021. These peaks coincided with Taiwan's local COVID-19 outbreak lasting from May to June 2021.

Conclusions: This study developed a blockchain-enabled mobile app, which can periodically retrieve and integrate PHRs from the NHIA MHB's cross-hospital data and the investigated hospital's self-pay medical data. Analysis of users' access logs revealed that the COVID-19 pandemic substantially increased individuals' use of PHRs and their health awareness with respect to COVID-19 prevention.

(J Med Internet Res 2022;24(1):e33399) doi:10.2196/33399

KEYWORDS

personal health records; COVID-19; My Health Bank; blockchain; public health

Introduction

Electronic medical records (EMRs) have enabled the digital transformation of health facilities worldwide [1-4]. With the widespread development and use of wearable devices, electronic health records (EHRs) can be generated and recorded effectively, and their adoption outside of the hospitals has increased rapidly [5-9]. Consequently, health data are expected to become more personalized and self-controlled after they are converted into personal health records (PHRs) [8-11]. With PHRs, patients can access, control, and track their health data and minimize hospital visits. Many countries have implemented national PHR systems that empower people to participate actively in their care and choose between opt-in and opt-out models [12-16]. However, despite the emerging trend of using PHRs, the PHR adoption rate remains low because of multiple challenges and barriers, including interoperability challenges relating to interfacility EMR and EHR access, implementation costs, barriers imposed by health care data security and privacy regulations (such as General Data Protection Regulation), and the assessment of relevant benefits for patients, health care providers, and health insurance institutes [7,8,17,18].

In 2014, Taiwan's National Health Insurance Administration (NHIA) launched the My Health Bank (MHB) 1.0 service [19]. Similar to other nation-based PHR systems [14-16], the MHB service provides a personal health account for each insured individual. Notably, the MHB continually collects insured individuals' medical data pertaining to outpatient visits, hospitalizations, dental services, traditional Chinese medicine outpatient visits, allergies, pathological test reports, X-ray, computed tomography, and magnetic resonance imaging examination results, discharge summaries, advance directives relating to organ donation and palliative care, vaccinations, and preventive care [20]. When an insured individual wants to retrieve his/her medical data, he/she can visit the MHB website, insert his/her National Health Insurance (NHI) smart card into a card reader, key in the corresponding password for the card, and download his/her MHB data as an XML, HTML, or JSON file after completing a verification process. In 2016, an updated version of the MHB, MHB 2.0, was launched. The new MHB incorporates an advanced encryption method that enables an insured individual to use his/her NHI smart card number and password to access his/her MHB data without having to use the physical card. In 2019, the NHIA released the MHB software development kit (SDK), which can be used to develop a mobile app for retrieving MHB data through an identity verification process.

The MHB SDK can attract more health care providers and startups to create innovative mobile apps in the health industry and encourage them to invest their resources into realizing market opportunities. It can also drive more users to engage in the control and management of their health data. However, the MHB SDK has several limitations that must be addressed. First, only the 3 most recent years of an individual's data are stored in the MHB. Therefore, a person cannot retrieve older medical data [16]. Second, the private medical or health data of self-pay patients, such as health checkup reports, are not stored in the MHB. Third, the personal MHB that an individual has downloaded is saved and stored in his/her mobile device. Hence, a data protection mechanism is required for such data given their sensitivity and confidentiality.

Because of the weak gatekeeper role of Taiwan's NHI system, patients can freely seek clinical services without referral whenever they feel uncomfortable [21]. Thus, they may not be aware of the benefits of using their PHRs. Moreover, NHI-insured individuals' participation in the NHIA MHB service is voluntary and not incentivized, and the corresponding participation rate thus remains low [19]. During the COVID-19 pandemic, the Taiwan government introduced a Name-Based Mask Distribution system that is linked to the NHIA MHB service. This system is a rationing system for face masks, which allows the public to purchase masks at a convenience store. Thus, the previously poor participation rates of the NHIA MHB service increased substantially during the COVID-19 pandemic [22]. In other words, people may have become aware of their medical data through the NHIA MHB service when they visited a hospital during the COVID-19 pandemic. Therefore, the COVID-19 pandemic may provide an opportunity for improving PHR use and accelerating the digital transformation of Taiwan's health care system.

In our previous study, we adopted the Go Ethereum version 1.7.3-stable to construct the iWellChain Framework, which is a permissioned consortium blockchain network with trusted parties to ensure consensus by proof of authority. Thus, the framework can limit participants who transact on the blockchain and define users who can serve the network by writing new blocks into the chain. In this way, the iWellChain Framework can assist the cooperative health care parties to establish their blockchain environment efficiently and further acquire the patient's EMRs and EHRs in accordance with the patient's signed Ethereum-based smart contracts. Accordingly, the framework could help health care parties to reduce their implementation costs and difficulties associated with the hospital's legacy information technology systems and security features surrounding data access. Moreover, the iWellChain

Framework provides a better secure data protection because there is no centralized structure for a malicious user to target, as the patients' EMRs and EHRs are stored in numerous copies on different blockchain nodes. With the iWellChain Framework, we also developed a blockchain-enabled mobile app, the iWellChain app, as a PHR tool for patients to acquire their EMRs and EHRs from the investigated hospital and cooperative clinics [23]. Launched on September 16, 2018, the iWellChain app is compatible with both iOS and Android devices. Thus, when a family doctor or specialist is interested in the referral patient's EMRs and EHRs, the patient can use the iWellChain app to authorize and set the approved period for the physician to access the EMRs and EHRs. However, it was a pilot study, and the limitations of the iWellChain app were that the patient could not use it to retrieve the patient's EMRs and EHRs from other health care facilities. Accordingly, our practical experience has indicated that iWellChain app is less popular than expected.

In this study, we aimed to integrate the MHB SDK with our developed blockchain-enabled PHR mobile app, which enables patients to access, store, and manage their cross-hospital PHR data. We also collected and analyzed the app's log data to examine patients' MHB use during the COVID-19 pandemic. This study involved the launch of an updated version of the iWellChain app in July 2019. The updated version incorporates the MHB SDK. Therefore, patients can regularly use it to retrieve, store, and track cross-hospital MHB data. Consequently, the iWellChain app addresses time restrictions related to the MHB data retention period. Additionally, the iWellChain app incorporates blockchain technology. Hence, a user can employ a public key to encrypt downloaded MHB data and ensure that their personal health data are secure. Furthermore, to encourage more individuals to use the iWellChain app, the investigated hospital released patients' self-pay medical data and health checkup reports, which are not available in the MHB database. Finally, to enhance our understanding of patients' PHR use during the COVID-19 pandemic, we analyzed the access logs of iWellChain app users to examine patients' activities during a 24-month study period.

Methods

Settings

This study was conducted at the Taipei Medical University Hospital (TMUH)—a teaching hospital with almost 800 beds and a good information infrastructure. In 2018, this hospital released five types of EMR and EHR that patients can access through the iWellChain app: outpatient department (OPD) notes, OPD laboratory test reports, discharge notes, pathology reports, and health check reports. An updated version of the iWellChain app was integrated with the MHB SDK and released in July 2019. This version enables patient access to emergency department (ED) notes and ED laboratory test reports.

Integration Workflow of the iWellChain App and MHB SDK

The MHB SDK was used to develop a mobile app that enables users to access their NHIA MHB data by completing an identity verification process. Figure 1 presents the integration workflow of the MHB SDK and mobile app. First, the mobile app user employs an application key acquired from and approved by the NHIA to initialize the MHB SDK. Thereafter, the mobile app invokes the MHB SDK to sign into the user's NHIA MHB account, and the user then completes an identity verification process that requires the user's national ID number, national insurance card number, password for the NHIA MHB service, and a randomized graph validation code (left panel of Figure 2). When the user's identity has been verified, he/she can select specific MHB data or all NHIA MHB data recorded within a specific time period. Next, the MHB SDK displays a window containing third-party statements relating to the use of NHIA MHB data (right panel of Figure 2). After the user agrees to the stated terms and conditions, the MHB SDK retrieves the user's NHIA MHB data in accordance with the set conditions and writes and stores the data to a log as a file ticket in the mobile app. Second, through the MHB SDK, the mobile app can use the file ticket to check the applicant's production status with respect to the NHIA MHB data on the NHIA server side. If the server process has not yet been completed, the MHB SDK then sends the mobile app a message stating that "No files can be downloaded." After the server process is complete, the mobile application then requests the applicant's MHB data for download, and the MHB SDK sends the message "The files are fully downloaded" and a unique server key to the mobile app. The NHIA MHB data are downloaded as a zip file that is encrypted using the Public-Key Cryptography Standards #5 password-based cryptography method defined in RFC 2898 [24]. Third, the mobile app uses the received application key as the password and the unique server key as the salt to generate a file key. Next, the mobile app uses the file key to extract the zip file and access the applicant's MHB data.

Figure 1. Integration workflow of the My Health Bank software development kit and mobile app. MHB: My Health Bank, NHIA: National Health Insurance Administration, SDK: software development kit.

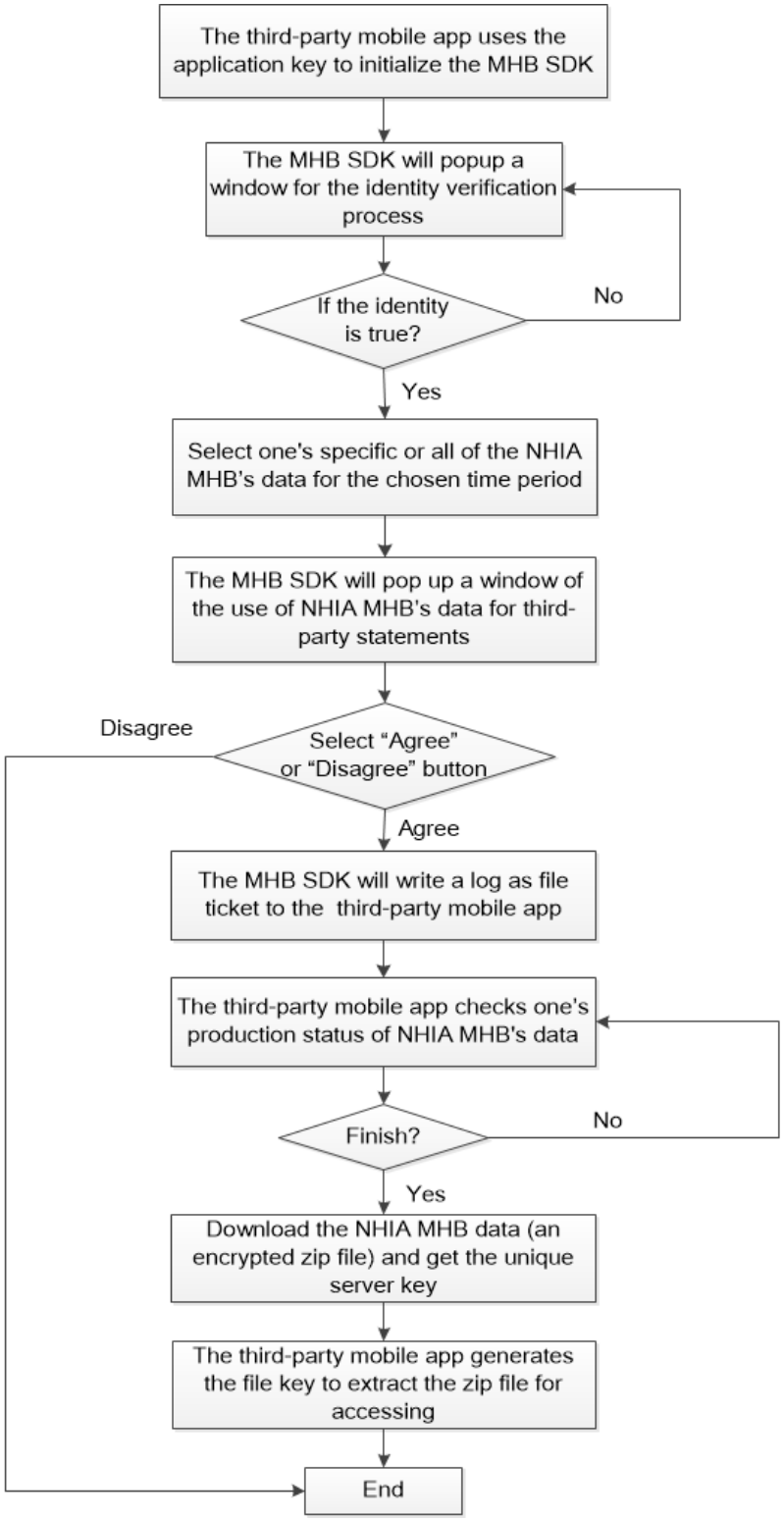
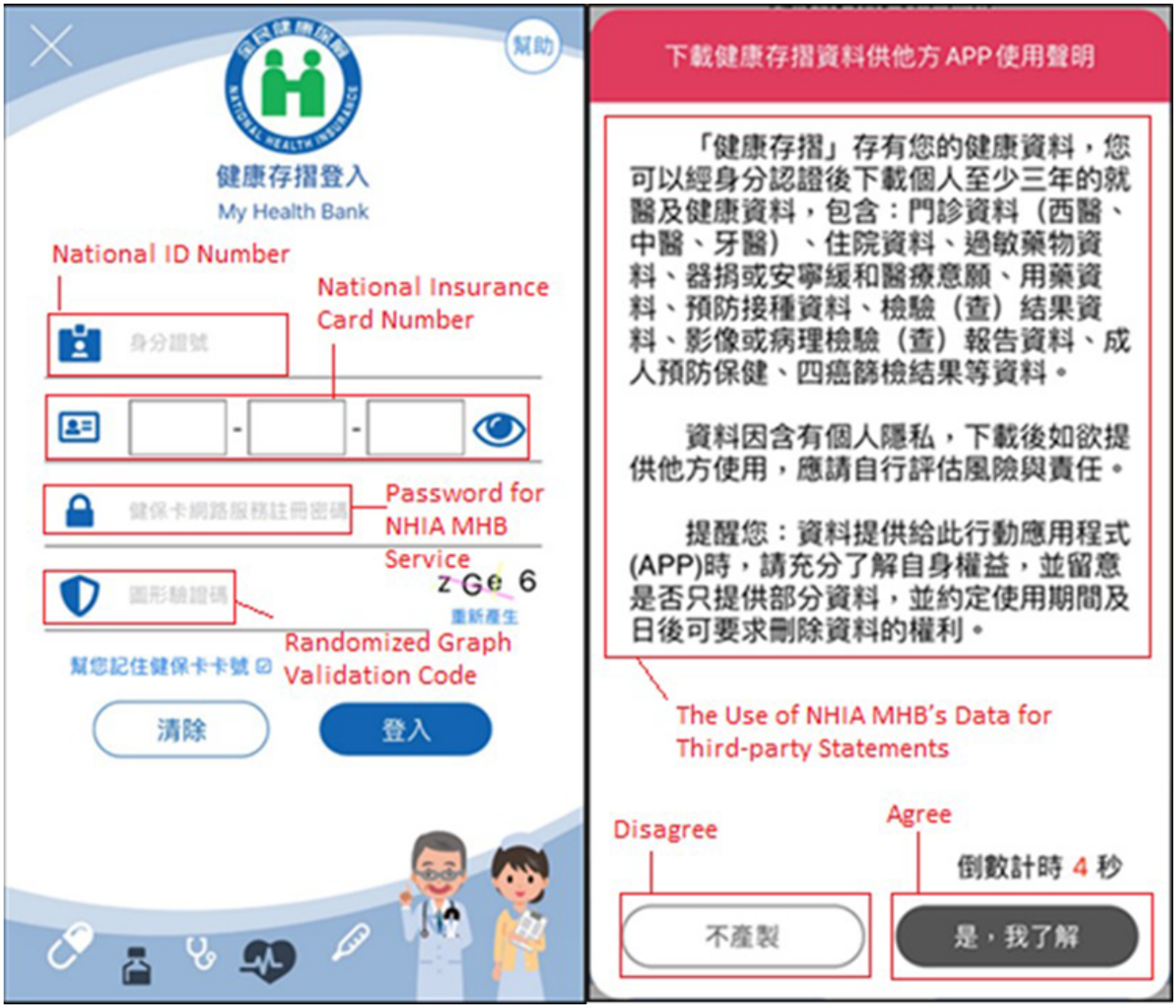


Figure 2. Identity verification process of the My Health Bank software development kit for accessing My Health Bank services (left panel); third-party statements relating to use of My Health Bank data (right panel).



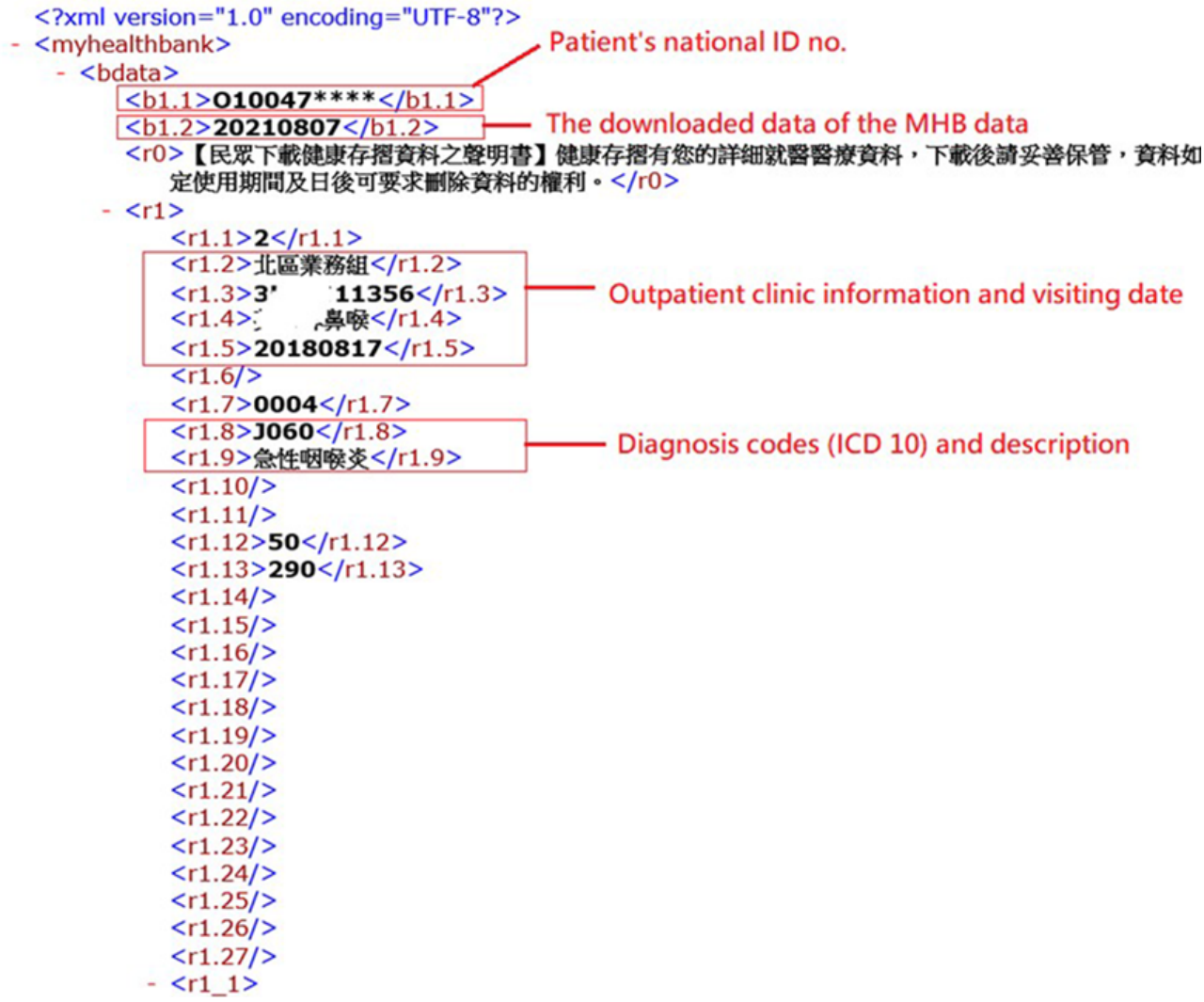
Analysis of NHIA MHB Data

The content of the downloaded NHIA MHB data is presented in a tagged format, such as XML or JASON format. Each tag contains a name, value, and the corresponding medical data. Table 1 presents the tag name and its corresponding medical data. Figure 3 presents the NHIA MHB data in the XML format. For example, the tag name "r1" represents outpatient clinic data, which include clinic information, the dates of visits, diagnosis codes, and descriptions. Accordingly, the iWellChain app extracts a patient's relevant medical data from the NHIA MHB data for further processing.

Table 1. Tag codes and values of National Health Insurance Administration My Health Bank data.

Name	Value
bdata	Patient information
r1	Outpatient clinic data
r2	Admission data
r3	Dental clinic data
r4	Allergy data
r5	Organ donation or palliative medicine
r6	Vaccination data
r7	Laboratory data
r8	Imaging or pathological examination reports
r9	Chinese medicine data
r10	Health and preventive care

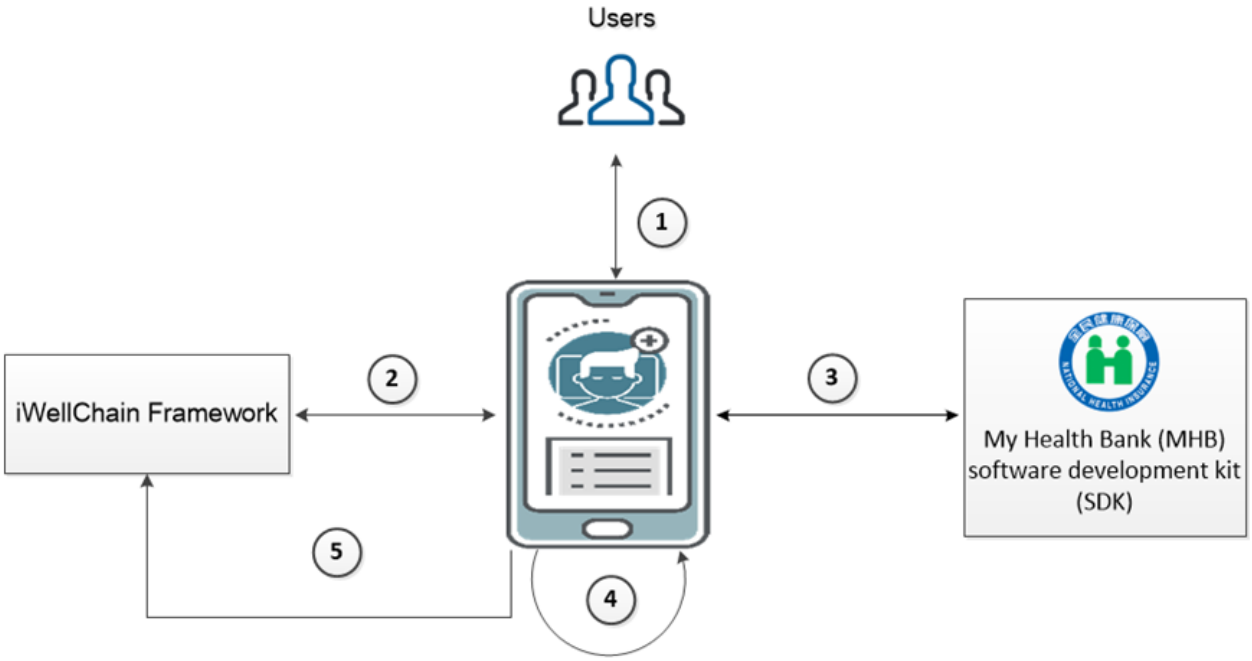
Figure 3. Downloaded My Health Bank data in XML format.



Data Processing and Synchronization Between iWellChain App and NHIA MHB Data

In our previous study, we constructed the iWellChain Framework—a permissioned consortium blockchain that utilizes trusted parties to ensure consensus by proof of authority [23]. Figure 4 describes the 5 steps of data processing and synchronization achieved through the iWellChain Framework between the iWellChain app and NHIA MHB data.

Figure 4. Interactions with the iWellChain app and My Health Bank data through the iWellChain Framework.



After a patient has created his/her blockchain account, he/she can use his/her national ID and application password and the password for the private key to sign into the iWellChain app.

Next, the iWellChain app synchronizes his/her Ethereum blockchain ledger through the iWellChain Framework. Thus, the patient can quickly obtain EMR and EHR indexes and further acquire the physical EMR and EHR files of these indexes using the iWellChain app.

After a patient has successfully signed into the NHIA MHB, his/her NHIA MHB data are downloaded. Once the NHIA MHB data are fully downloaded, the iWellChain app decrypts these data and stores them in the mobile app sandbox.

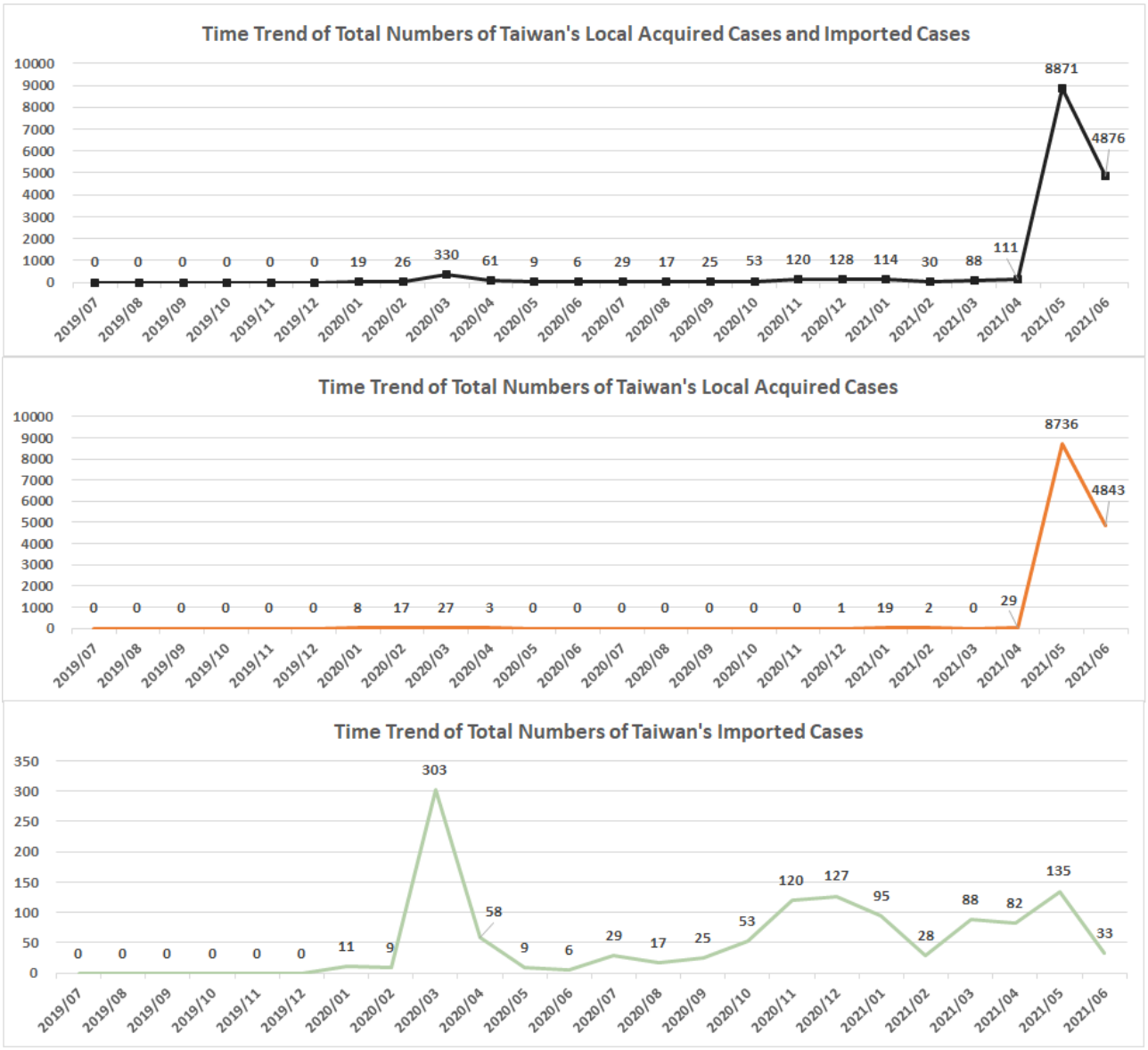
The iWellChain app then analyzes each of the patient's NHIA MHB data records to verify that these records were not previously acquired or did not originate from other health facilities.

For each record that has been verified as a new record, the iWellChain app synchronizes with the iWellChain Framework to request the patient's public key, which is then used to encrypt the record's path. Finally, the iWellChain app registers the record's information and its encrypted path in the iWellChain Framework.

Trends in the COVID-19 Pandemic in Taiwan

According to Taiwan's Central Epidemic Command Center (CECC) reports, confirmed COVID-19 cases can be classified as locally acquired cases (LACs) and imported cases (ICs). Figure 5 indicates that both LACs and ICs started arising in January 2020. From January to April 2020, both LACs and ICs were reported: 19 (8 LACs, 11 ICs), 26 (17 LACs, 9 ICs), 330 (27 LACs, 303 ICs), and 61 (3 LACs, 58 ICs) cases were reported in January, February, March, and April 2020, respectively. During the subsequent 7-month period (May 2020 to November 2020), no LACs were reported. However, the number of ICs reported fluctuated from month to month. During the next 3-month period (December 2020 to February 2021), both LACs and ICs were reported: 128 (1 LAC, 127 ICs), 114 (19 LACs, 95 ICs), and 30 (2 LACs, 28 ICs) cases were reported in December 2020, January 2021, and February 2021, respectively. In March 2021, 0 LACs and 88 ICs were reported. From April to June 2021, the number of LACs reported increased substantially: 29 LACs and 82 ICs, 8,734 LACs and 135 ICs, and 4,785 LACs and 33 ICs were reported in April, May, and June 2021, respectively. Accordingly, the CECC raised the COVID-19 alert level to level 3 nationwide and maintained it from May 19 to July 26, 2021. Per COVID-19 level 3 alert guidelines, places of business and public venues were recommended to stop operating (excluding essential services, law enforcement, medical treatment facilities, and government offices). However, for places that remained open, strict mask-wearing and social-distancing protocols were enforced [25].

Figure 5. Taiwan's locally acquired and imported case trends.



Log Data Collection and Analysis of the iWellChain App

Because of the COVID-19 pandemic, the investigated hospital had to adhere to CECC guidelines regarding the reservation of general wards for the influx of patients with COVID-19. Accordingly, in this study, we only included patients attending the OPD and ED of TMUH and used the iWellChain app. We retrospectively collected patients' iWellChain app access logs over a 24-month period (before and during the COVID-19 pandemic) from July 2019 to June 2021. Accordingly, an access log database was created, and information on the actions that the patients performed through the iWellChain app was collected. We integrated Google Analytics for Firebase into the iWellChain app to capture patients' click data. Once the data had been captured and stored, the access log could be linked to the Google Colaboratory [26]—a web-based and free Python development platform for further analysis and reporting. The Joint Institutional Review Board of Taipei Medical University and TMUH approved this study.

Results

PHR Features of the iWellChain App

When a patient signs into his/her blockchain account, the iWellChain app displays his/her subscribed PHR services per the signed Ethereum-based smart contracts. Figure 6 (left panel) presents an example of a patient who subscribed to three PHR services: the employee card service (top section), NHIA MHB card service (middle section), and TMUH member card service (bottom section). The employee card service is designed to enable employees to access their annual regular checkup reports. This service can also be used to establish an employee's personal health profile for long-term health tracking and management. The NHIA MHB card service can activate the MHB SDK to retrieve an individual's MHB data after he/she has completed the identity verification process. The TMUH member card service provides a single window in which the patient can acquire and integrate EMR, EHR, and NHIA MHB data as the PHR data. Figure 6 (right panel) presents the timeline of a patient's medical data that were sorted in accordance with his/her

visit dates (from latest to oldest). For example, the patient had three OPD notes and one laboratory test report. The first OPD note was issued by the investigated hospital on January 11, 2021. The next two OPD notes and one laboratory test report were retrieved from the patient's NHIA MHB data. The two OPD notes were issued by dental clinic A and hospital B on December 5, 2020, and the laboratory test report was issued by hospital C on March 17, 2020. The patient could also select the specific medical data to view more details. Figure 7 presents the patient's prescription information, which includes names of medical orders, total medication dosages, and days of medication administration.

Figure 6. Patient's subscribed personal health record services in the iWellChain app (left panel) and medical data sorted by visit date from latest to oldest.

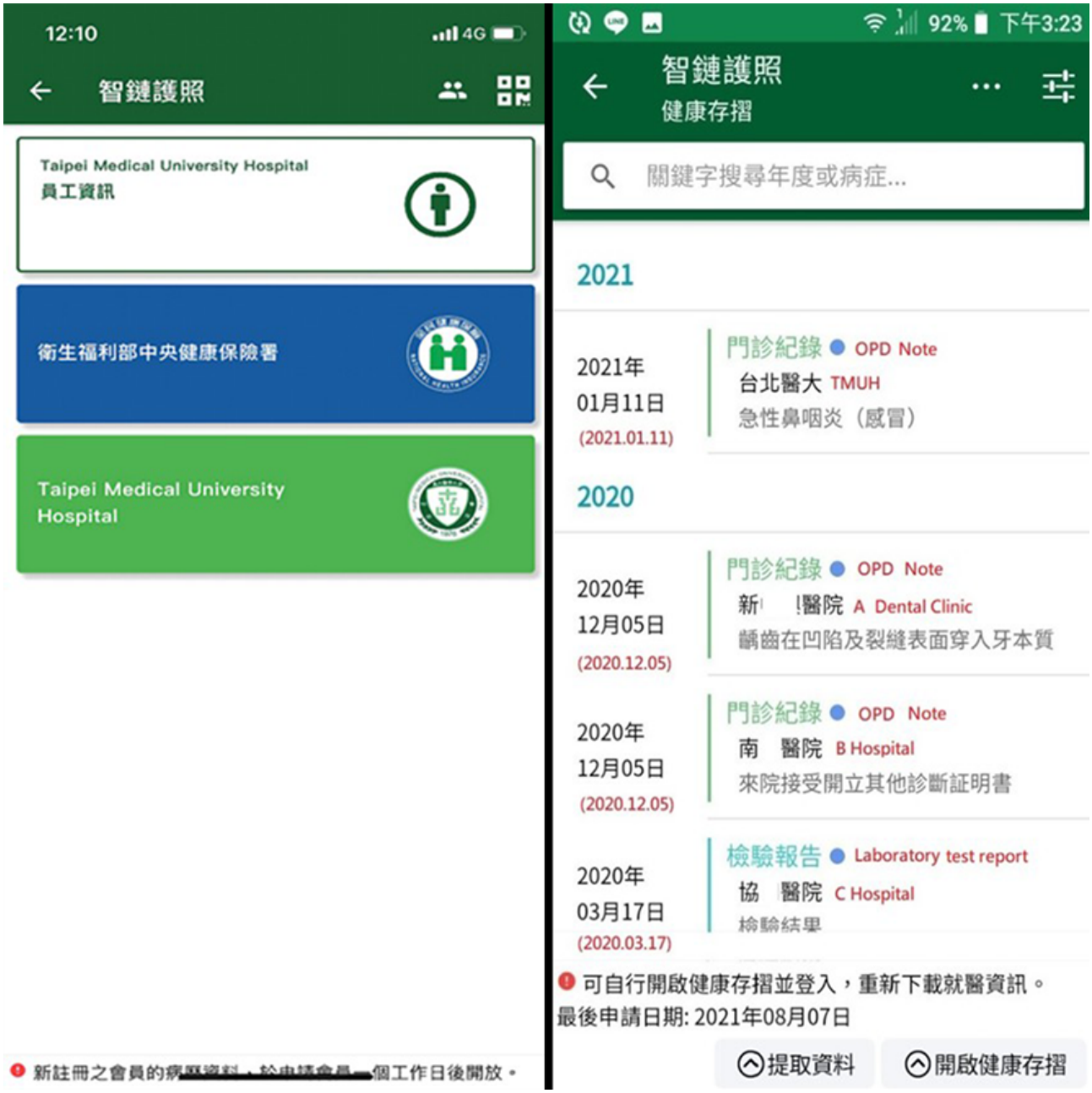


Figure 7. Patient's outpatient department notes and corresponding prescription information.

Study Population

We collected 24 months of data (July 1, 2019, to June 30, 2021) from the investigated hospital. During this 2-year period, the total number of patients who visited the OPD and ED was 1,405,316, and 92,184, respectively. From July 1, 2019, to December 30, 2019, the proportion of patients who visited the ED and OPD who subscribed to the iWellChain app ranged from 0.29% to 1.67% and from 3.09% to 5.36%, respectively. In 2020, the proportion of patients who visited the ED and OPD who subscribed to the iWellChain app ranged from 1.41% to 2.54% and from 5.29% to 6.25%, respectively. Finally, from January 1, 2021, to June 30, 2021, the proportion of patients who visited the ED and OPD who subscribed to the iWellChain app ranged from 1.95% to 8.20% and from 5.58% to 7.49%, respectively. Therefore, the iWellChain subscription trend for patients who visited the OPD or ED was a steady increase during the investigated period. The descriptive statistics for the study population are presented in Table 2.

Table 2. Proportion of patients who visited the outpatient department and emergency department and who subscribed to the mobile app between July 2019 and June 2021.

Month#	ED[a] patients, n	ED patients subscribed to the app, n (%)	OPD[b] patients, n	OPD patient subscribed to the app, n (%)
2019				
7	4199	12 (0.29)	65,152	2011 (3.09)
8	4194	22 (0.52)	62,841	2178 (3.47)
9	4266	27 (0.63)	61,108	2354 (3.85)
10	4250	17 (0.4)	64,454	2544 (3.95)
11	3877	32 (0.83)	64,080	2676 (4.18)
12	4187	70 (1.67)	63,709	3413 (5.36)
2020				
1	4980	89 (1.79)	59,506	3150 (5.29)
2	3421	71 (2.08)	53,918	2919 (5.41)
3	3226	68 (2.11)	48,970	3052 (6.23)
4	2842	51 (1.79)	47,116	2833 (6.01)
5	3186	45 (1.41)	54,119	3185 (5.89)
6	3427	76 (2.22)	54,100	3308 (6.11)
7	3549	90 (2.54)	63,144	3829 (6.06)
8	3577	81 (2.26)	61,902	3607 (5.83)
9	3647	72 (1.97)	63,712	3680 (5.78)
10	4040	71 (1.76)	65,443	3667 (5.6)
11	3788	74 (1.95)	61,820	3707 (6)
12	3546	74 (2.09)	63,752	3986 (6.25)
2021				
1	3835	79 (2.06)	59,872	3530 (5.9)
2	3602	89 (2.47)	52,761	2942 (5.58)
3	3995	78 (1.95)	65,574	4083 (6.23)
4	3829	82 (2.14)	64,043	3812 (5.95)
5	5025	412 (8.2)	46,908	3394 (7.24)
6	3696	208 (5.63)	37,312	2794 (7.49)

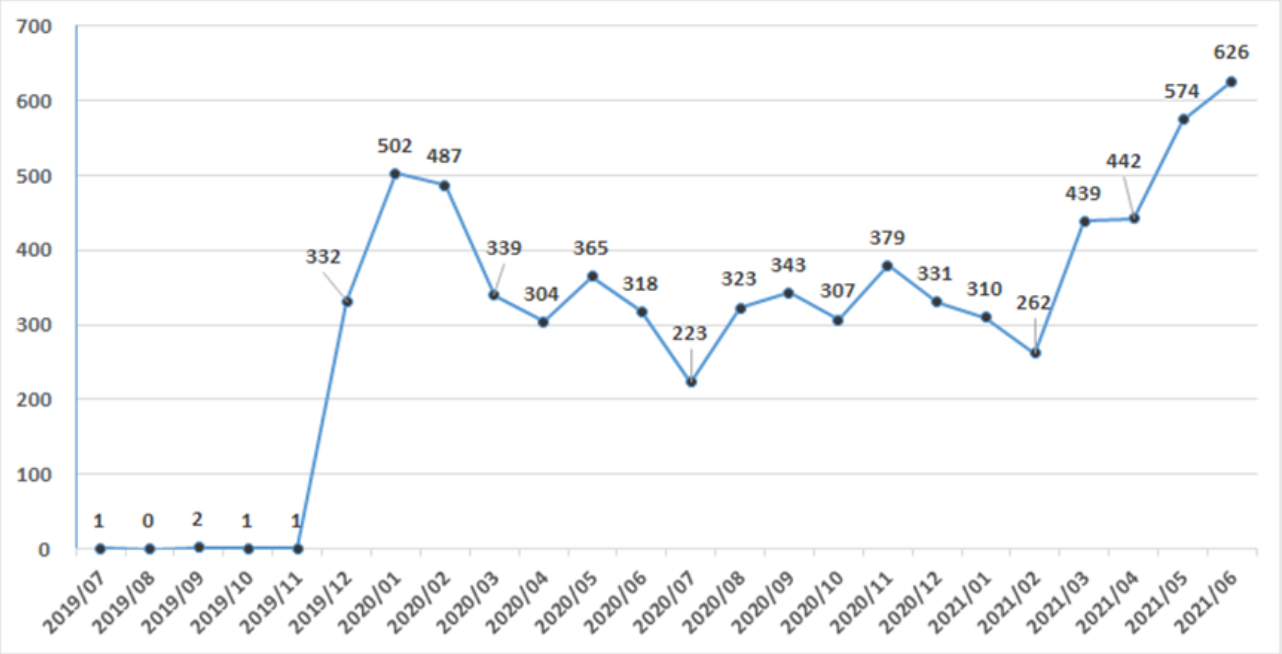
[a]ED: emergency department.

[b]OPD: outpatient department.

New Blockchain Account Applicants

The iWellChain app has several basic functions (ie, an appointment service, inquiring for a doctor's information, and a payment service); however, patients have to apply for and register web-based blockchain accounts to acquire their medical data. Figure 8 presents the trend in the numbers of new applicants. During the study period, the total number of patients who applied for a blockchain account was 7211. From July to December 2019, only a few new people applied; specifically, the number of new applicants was 1, 0, 2, 1, 1, and 332 in July, August, September, October, November, and December 2019, respectively. Since January 2020, the number of new applicants increased substantially, and the number of applicants was 6874 (95.33%) from January 2020 to June 2021. The number of applicants peaked at 626 in June 2021; the second (574) and third (502) highest numbers were observed in May 2021 and January 2020, respectively. The peak in the number of new applicants coincided with the onset of COVID-19 and outbreak of LACs in Taiwan.

Figure 8. Numbers of new applicants of the blockchain accounts.

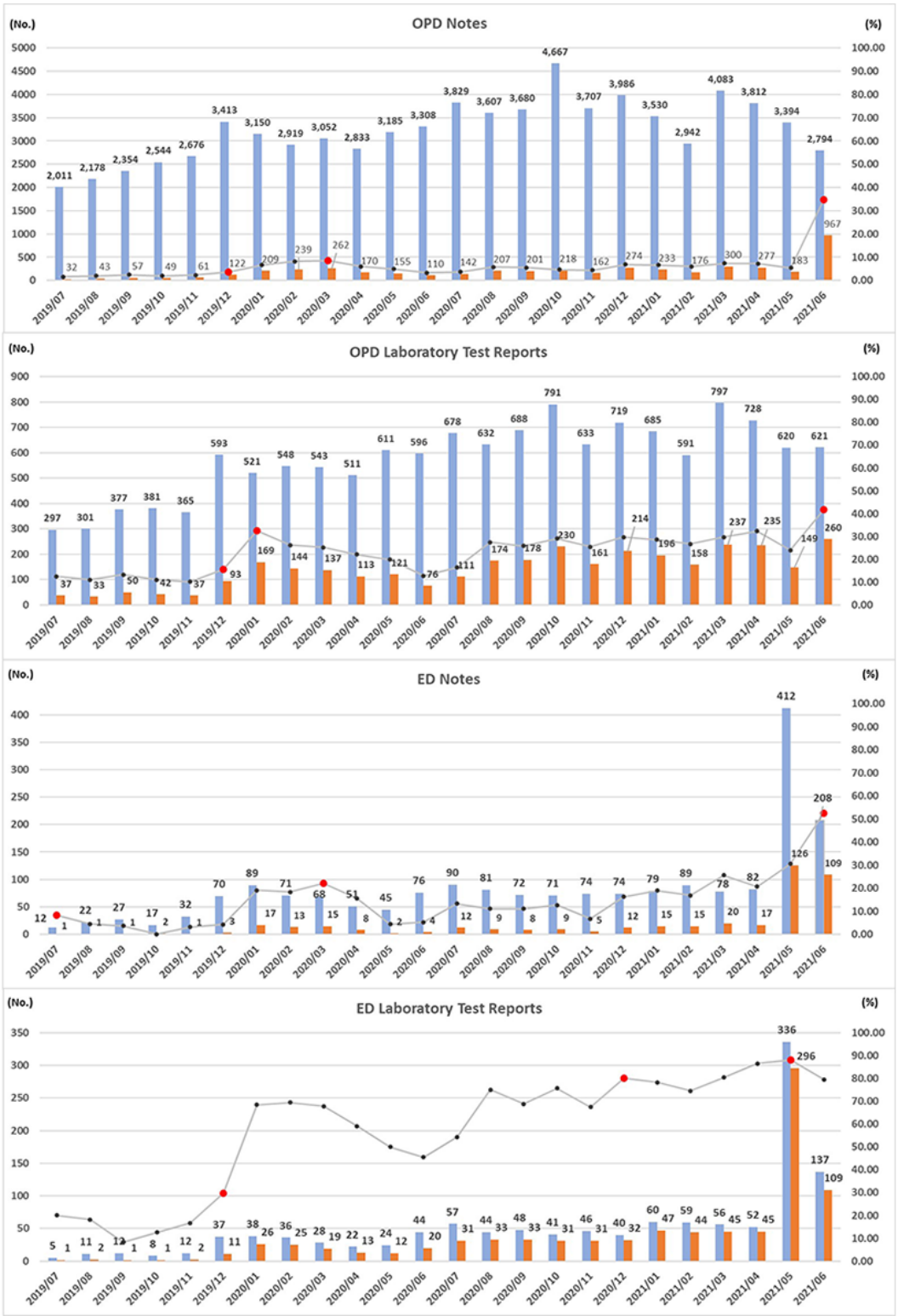


Analysis of Patients' Access Log Data for Medical Data Acquisition

Through the access log data, we further examined the investigated hospital's data set of patients who visited the OPD and ED, with data including OPD notes, OPD laboratory test results, ED notes, and ED laboratory test reports. The content of the OPD and ED notes include OPD or ED visit dates, chief complaints, diagnosis codes and descriptions, and prescription information.

In Figure 9, the left axis indicates that the numbers of patients visiting the OPD and ED who subscribed to the iWellChain app and had visited the hospital (A) and the numbers of patients who used the iWellChain app to acquire the corresponding OPD notes, OPD laboratory test results, ED notes, and ED laboratory test reports (B). The right axis represents the value obtained by dividing value (B) by value (A) (B/A). The number (A) and proportion (B/A) of patients who visited the OPD and ED for each month varied depending on the type of medical data.

Figure 9. Analysis of patients' access log data with respect to acquisition of outpatient department (OPD) notes, OPD laboratory test results, emergency department (ED) notes, and ED laboratory test reports through the iWellChain app.



During the 2-year study period, subscribers who retrieved OPD or ED notes and laboratory test reports accounted for the highest proportion of patients who visited the OPD or ED in May and June 2021. These peak proportions were 52.40%, 88.10%, 70.30%, 34.61%, 41.87%, and 59.85% for OPD notes (June 2021), OPD laboratory test results (June 2021), ED notes (June 2021), and ED laboratory test reports (May 2021), respectively.

From July 2019 to December 2019, only a few patients visiting the OPD or ED subscribed to the iWellChain app and used it to acquire their OPD or ED notes and laboratory test reports (Figure 9). For example, 2011, 2178, 2354, 2676, 2715, and 3414 patients who visited the OPD subscribed to the application in July, August, September, October, November, and December 2019, respectively. Among these patients, only 32 (1.59%), 43 (1.97%), 57 (2.42%), 49 (1.93%), 61 (2.28%), and 122 (3.57%)

used the iWellChain app to acquire their OPD notes in July to December 2019, respectively. Furthermore, only some patients who visited the OPD underwent laboratory tests. Among the patients who visited the OPD who subscribed to the app in July, August, September, October, November, and December 2019, only 297, 301, 377, 381, 365, and 593, respectively, underwent laboratory tests. Among them, only 37 (12.46%), 33 (10.96%), 50 (13.26%), 42 (11.02%), 37 (10.14%), and 93 (15.68%) patients used the iWellChain app to acquire their OPD laboratory test reports, respectively.

In 2019, for each type of medical data, the peak proportions were thus 8.33% for ED notes (July 2019), 29.73% for ED laboratory test reports (December 2019), 3.57% for OPD notes (December 2019), and 15.68% for OPD laboratory test reports (December 2019). In 2020, the peak proportions for each type of medical data were all higher than those observed in 2019. Moreover, from January to June 2021, the peak proportions for each type of medical data and report were all higher than those observed in 2019 and 2020.

Discussion

Principal Findings

Taiwan's NHI system is a single-payer system that was introduced in 1995; by 2010, it was already providing universal and mandatory insurance coverage for almost all of Taiwan's population (99.5%) [27,28]. The continual accumulation of NHI MHB data provides a favorable opportunity for using the data as a catalyst to improve PHR use through the MHB SDK. Utilization of the MHB SDK to implement a PHR app is a promising implementation method because of the low implementation cost, high data quality, and availability without participation bias [28].

In this study, we successfully integrated the MHB SDK into our blockchain-enabled mobile app to enable use by patients. The iWellChain app's implementation has three advantages. First, iWellChain app can help patients periodically acquire their own cross-hospital EMRs and avoid the time restrictions relating to the NHIA MHB data retention period. Second, to protect patients' privacy and data security, the downloaded NHIA MHB data are stored on the patients' mobile devices without moving these data to or depositing them in other storage or cloud-based spaces. With our method, patients feel more secure in using the iWellChain app for PHR data management. Third, the investigated hospital also released self-pay patients' EMRs and EHRs (eg, health checkup reports) to complement the deficiencies of NHIA MHB data. However, under the Taiwan NHI system, people could easily obtain in-person clinic/hospital medical consultations with relatively low costs. Consequently, people would rather go to the hospital or clinic than use such the PHR mobile app to acquire their medical or health information. A previous study indicated the NHI-insured individuals' participation rate in the NHIA MHB service was low [19]. Moreover, in this study, although the updated iWellChain app combined with the MHB SDK launched, it appeared not to attract more patients to use it for acquiring their PHRs in the beginning period (from July 2019 to December 2019). Accordingly, providing effective PHR mobile apps might be not the critical factor to increase the app users.

Before May 2021, Taiwan only had a few LACs (Figure 5). Thus, before May 2021, the willingness of Taiwanese people to be vaccinated was low, widespread screening—such as COVID-19 polymerase chain reaction (PCR) or rapid tests—was not urgently required, and few people used PHR apps. In May 2021, a COVID-19 outbreak occurred in tea houses located in Taipei City's Wanhua District [29]. This outbreak quickly led to the announcement of a nationwide level 3 COVID-19 alert. Accordingly, the CECC established rapid testing sites in multiple cities, with the focus being Taipei City and New Taipei City. When an individual feels uncomfortable, he/she can go to a rapid testing site or a hospital to undergo COVID-19 rapid or PCR testing. After the test result has been released, the rapid testing site or hospital uploads the result to the NHIA database, which then synchronizes with the database used by the Taiwan Centers for Disease Control (CDC). However, the heavy load of mass screenings meant that individuals might have to wait between 3 and 6 days before they could obtain their test result and report on the CDC's website or NHIA MHB's database. Concurrently, the trends observed in the investigated hospital (located in Taipei City) reflected the general trend in Taiwan's COVID-19 outbreak. Compared with the government's COVID-19 test reports, the iWellChain app reported the results of COVID-19 rapid and PCR tests within 1 day for patients who visited the OPD or ED of the investigated hospital. Figure 8 indicates that more people applied to create a blockchain account in May and June 2021 relative to other months. Table 2 indicates that the proportion of patients who visited the ED and OPD who subscribed to the app peaked at 8.20% (May 2021) and 7.49% (June 2021), respectively; among the aforementioned patients, the proportions who used the app to acquire their ED and OPD notes (peaks in May 2021) and laboratory test reports (ED peak in May 2021; OPD peak in June 2021) also peaked during the same period. These findings suggest that the COVID-19 pandemic led patients to pay more attention to their medical data. They also reflect a substantial increase in patients' use of the PHR app in response to Taiwan's local COVID-19 outbreak.

Under Taiwan's NHI system, insured individuals' medical data are routinely collected through a network with well-developed infrastructure. Although Taiwan has been affected by the COVID-19 outbreak, the CECC could effectively coordinate with all levels of health care facilities to control and manage epidemic information in response to the disease's spread (eg, establishment of a rapid testing station in a hotspot or contact tracing) by implementing the existing infectious disease notification method. Taiwan lifted the Level 3 alert on July 27, 2021, and only a few LACs were reported in August 2021 [30]. Furthermore, the NHIA MHB database is being continually expanded with respect to vaccination information and progress. Similar to the bundling of the NHIA MHB service with mask purchase services, our investigated hospital provided the public with a web-based service for booking an appointment for a leftover vaccine through the iWellChain app [31]. This model could attract more users to use our iWellChain app.

Conclusions

In a patient-centered model, EMRs and EHRs belong to the individual patient. Furthermore, Taiwan's NHIA MHB database offers a robust foundation for PHR development. This study provides a blockchain-enabled mobile app that can periodically retrieve and integrate cross-hospital PHRs from the NHIA MHB database and the investigated hospital's self-pay medical data and provide secure data protection through blockchain technology. The user access log analysis indicates that the COVID-19 pandemic has had a substantial effect on the app's use, increasing individuals' PHR use and health awareness regarding COVID-19 prevention. However, compared with the investigated hospital's total number of patients who visited the OPD and ED, the number of app users remains low. Therefore, use of the iWellChain app for PHR acquisition can be further improved.

Acknowledgments

We would like to thank all information team members and the Preventive and Community Medicine Department of TMUH for their support. We would also like to thank and recognize the Radica Health startup for their technical contributions. This study was supported in part by the Ministry of Science and Technology of Taiwan (grant MOST 107-2823-8-038-001, 109-2221-E-305-006, 110-2221-E-305-002, 110-2218-E-305-001-MBK) and University System of Taipei Joint Research Program (grant TMU 104-AE1-B31, USTP-NTPU-TMU-109-02).

Conflicts of Interest

None declared.

References

1. Spencer K, Sanders C, Whitley EA, Lund D, Kaye J, Dixon WG. Patient Perspectives on Sharing Anonymized Personal Health Data Using a Digital System for Dynamic Consent and Research Feedback: A Qualitative Study. J Med Internet Res 2016 Apr 15;18(4):e66 [FREE Full text] [doi: 10.2196/jmir.5011] [Medline: 27083521]
2. Schofield P, Shaw T, Pascoe M. Toward Comprehensive Patient-Centric Care by Integrating Digital Health Technology With Direct Clinical Contact in Australia. J Med Internet Res 2019 Jun 04;21(6):e12382 [FREE Full text] [doi: 10.2196/12382] [Medline: 31165713]
3. Boeldt DL, Wineinger NE, Waalen J, Gollamudi S, Grossberg A, Steinhubl SR, et al. How Consumers and Physicians View New Medical Technology: Comparative Survey. J Med Internet Res 2015 Sep 14;17(9):e215 [FREE Full text] [doi: 10.2196/jmir.4456] [Medline: 26369254]
4. Pohlmann S, Kunz A, Ose D, Winkler EC, Brandner A, Poss-Doering R, et al. Digitalizing Health Services by Implementing a Personal Electronic Health Record in Germany: Qualitative Analysis of Fundamental Prerequisites From the Perspective of Selected Experts. J Med Internet Res 2020 Jan 29;22(1):e15102 [FREE Full text] [doi: 10.2196/15102] [Medline: 32012060]
5. Ryu B, Kim N, Heo E, Yoo S, Lee K, Hwang H, et al. Impact of an Electronic Health Record-Integrated Personal Health Record on Patient Participation in Health Care: Development and Randomized Controlled Trial of MyHealthKeeper. J Med Internet Res 2017 Dec 07;19(12):e401 [FREE Full text] [doi: 10.2196/jmir.8867] [Medline: 29217503]
6. Zheng X, Sun S, Mukkamala RR, Vatrapu R, Ordieres-Meré J. Accelerating Health Data Sharing: A Solution Based on the Internet of Things and Distributed Ledger Technologies. J Med Internet Res 2019 Jun 06;21(6):e13583 [FREE Full text] [doi: 10.2196/13583] [Medline: 31172963]
7. Dinh-Le C, Chuang R, Chokshi S, Mann D. Wearable Health Technology and Electronic Health Record Integration: Scoping Review and Future Directions. JMIR Mhealth Uhealth 2019 Sep 11;7(9):e12861 [FREE Full text] [doi: 10.2196/12861] [Medline: 31512582]
8. Roehrs A, da Costa CA, Righi RDR, de Oliveira KSF. Personal Health Records: A Systematic Literature Review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
9. Gay V, Leijdekkers P. Bringing Health and Fitness Data Together for Connected Health Care: Mobile Apps as Enablers of Interoperability. J Med Internet Res 2015 Nov 18;17(11):e260 [FREE Full text] [doi: 10.2196/jmir.5094] [Medline: 26581920]
10. de Jong JM, Ogink PA, van Bunningen CG, Driessen RJ, Engelen LJ, Heeren B, et al. A Cloud-Based Virtual Outpatient Clinic for Patient-Centered Care: Proof-of-Concept Study. J Med Internet Res 2018 Sep 24;20(9):e10135 [FREE Full text] [doi: 10.2196/10135] [Medline: 30249584]
11. Schleimer E, Pearce J, Barnecut A, Rowles W, Lizee A, Klein A, et al. A Precision Medicine Tool for Patients With Multiple Sclerosis (the Open MS BioScreen): Human-Centered Design and Development. J Med Internet Res 2020 Jul 06;22(7):e15605 [FREE Full text] [doi: 10.2196/15605] [Medline: 32628124]
12. Daruwalla ZJ, Wong KL, Thambiah J. The application of telemedicine in orthopedic surgery in singapore: a pilot study on a secure, mobile telehealth application and messaging platform. JMIR Mhealth Uhealth 2014 Jun 05;2(2):e28 [FREE Full text] [doi: 10.2196/mhealth.3303] [Medline: 25100283]

13. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors Affecting Patients' Use of Electronic Personal Health Records in England: Cross-Sectional Study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
14. Laranjo L, Rodolfo I, Pereira AM, de Sá AB. Characteristics of Innovators Adopting a National Personal Health Record in Portugal: Cross-Sectional Study. JMIR Med Inform 2017 Oct 11;5(4):e37 [FREE Full text] [doi: 10.2196/medinform.7887] [Medline: 29021125]
15. Bidargaddi N, van Kasteren Y, Musiat P, Kidd M. Developing a Third-Party Analytics Application Using Australia's National Personal Health Records System: Case Study. JMIR Med Inform 2018 Apr 24;6(2):e28 [FREE Full text] [doi: 10.2196/medinform.7710] [Medline: 29691211]
16. Yeh M, Saltman RB. Creating online personal medical accounts: Recent experience in two developed countries. Health Policy Technol 2019 Jun;8(2):171-178. [doi: 10.1016/j.hlpt.2019.05.004]
17. Abdolkhani R, Gray K, Borda A, De Souza R. Patient-Generated Health Data Quality for Clinical Use: Human and Technology Factors. iproc 2018 Sep 17;4(2):e11703. [doi: 10.2196/11703]
18. Kirkendall ES, Ni Y, Lingren T, Leonard M, Hall ES, Melton K. Data Challenges With Real-Time Safety Event Detection And Clinical Decision Support. J Med Internet Res 2019 May 22;21(5):e13047 [FREE Full text] [doi: 10.2196/13047] [Medline: 31120022]
19. My Health Bank. National Health Insurance Administration: Ministry of Health and Welfare. URL: https://www.nhi.gov.tw/english/Content_List.aspx?n=21D194F3C675DB0E [accessed 2021-04-10]
20. My Health Bank. URL: https://myhealthbank.nhi.gov.tw/IHKE0002/IHKE0002S07.aspx [accessed 2021-02-10]
21. Nugent T, Upton D, Cimpoesu M. Improving data transparency in clinical trials using blockchain smart contracts. F1000Res 2016;5:2541 [FREE Full text] [doi: 10.12688/f1000research.9756.1] [Medline: 28357041]
22. Tsai W, Wu Y, Cheng C, Kuo M, Chang Y, Hu F, et al. A Technology Acceptance Model for Deploying Masks to Combat the COVID-19 Pandemic in Taiwan (My Health Bank): Web-Based Cross-sectional Survey Study. J Med Internet Res 2021 Apr 21;23(4):e27069 [FREE Full text] [doi: 10.2196/27069] [Medline: 33819168]
23. Lo Y, Yang C, Chien H, Chang S, Lu C, Chen R. Blockchain-Enabled iWellChain Framework Integration With the National Medical Referral System: Development and Usability Study. J Med Internet Res 2019 Dec 04;21(12):e13563 [FREE Full text] [doi: 10.2196/13563] [Medline: 31799935]
24. PKCS# 5: Password-Based Cryptography Specification Version 2. Kaliski B. URL: https://datatracker.ietf.org/doc/html/rfc2898 [accessed 2020-11-11]
25. Chiang H, Chen C, Joseph Y. Level 3 COVID-19 Alert Introduced Across Taiwan. Central News Agency. URL: https://focustaiwan.tw/society/202105190017 [accessed 2021-05-20]
26. Welcome to Colab. Google Research. URL: https://colab.research.google.com/notebooks/intro.ipynb?hl=en [accessed 2021-07-02]
27. Sung S, Hsieh C, Hu Y. Two Decades of Research Using Taiwan's National Health Insurance Claims Data: Bibliometric and Text Mining Analysis on PubMed. J Med Internet Res 2020 Jun 16;22(6):e18457 [FREE Full text] [doi: 10.2196/18457] [Medline: 32543443]
28. Wu T, Majeed A, Kuo KN. An overview of the healthcare system in Taiwan. London J Prim Care (Abingdon) 2010 Dec;3(2):115-119 [FREE Full text] [doi: 10.1080/17571472.2010.11493315] [Medline: 25949636]
29. Zennie M, Tsai G. How a False Sense of Security, and a Little Secret Tea, Broke Down Taiwan's COVID-19 Defenses. Time. 2021 May 21. URL: https://time.com/6050316/taiwan-covid-19-outbreak-tea [accessed 2021-05-22]
30. Everington K. Taiwan to lower restrictions to Level 2 on July 27. Taiwan News. 2021 Jul 23. URL: https://www.taiwannews.com.tw/en/news/4255807 [accessed 2021-07-27]
31. Everington K. Taiwan CECC mulling opening up 'vaccine leftovers' to people over 30. Taiwan News. 2021 Jun 23. URL: https://www.taiwannews.com.tw/en/news/4230676 [accessed 2021-06-23]

Abbreviations

CDC: Centers for Disease Control
CECC: Central Epidemic Command Center
ED: emergency department
EHR: electronic health record
EMR: electronic medical record
IC: imported case
LAC: locally acquired case
MHB: My Health Bank
NHI: National Health Insurance
NHIA: National Health Insurance Administration
OPD: outpatient department
PCR: polymerase chain reaction

PHR: personal health record
SDK: software development kit
TMUH: Taipei Medical University Hospital

Edited by C Basch; submitted 14.09.21; peer-reviewed by CY Huang, I Mircheva; comments to author 20.11.21; revised version received 08.12.21; accepted 16.12.21; published 06.01.22.

Please cite as:
Tseng CH, Chen RJ, Tsai SY, Wu TR, Tsaur WJ, Chiu HW, Yang CY, Lo YS
Exploring the COVID-19 Pandemic as a Catalyst for Behavior Change Among Patient Health Record App Users in Taiwan: Development and Usability Study
J Med Internet Res 2022;24(1):e33399
URL: https://www.jmir.org/2022/1/e33399
doi:10.2196/33399
PMID:34951863

Original Paper

Early Adopters of Apple Health Records at a Large Academic Medical Center: Cross-sectional Survey of Users

Joshua Rolnick[1,2], MS, MD, JD; Robin Ward[3], MS; Gordon Tait[3], BS; Neha Patel[2,3], MS, MD

[1]United States Department of Veterans Affairs, Philadelphia, PA, United States

[2]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

[3]University of Pennsylvania Health System, Philadelphia, PA, United States

Corresponding Author:
Joshua Rolnick, MS, MD, JD
United States Department of Veterans Affairs
3900 Woodland Avenue
Philadelphia, PA, 19104
United States
Phone: 1 6175385191
Email: rolnick@pennmedicine.upenn.edu

Abstract

Background: Mobile applications offer a new approach to personal health records, which are internet-based tools for patients to consolidate and manage their health information. The University of Pennsylvania Health System (UPHS) was one of the first health systems to participate in Apple Health Records (AHR), a prominent example of this new generation of personal health records.

Objective: This study aimed to characterize early adoption of AHR among UPHS patients and understand user perspectives.

Methods: An email-based survey with fixed answer, Likert scale, and open-ended questions was administered to all UPHS patients using AHR in the first 10 months of enrollment. Survey data linked to the UPHS electronic health record system were used to analyze responses. Multivariable logistic regression modeled the association of patient characteristics with user ratings. Content analysis was used to analyze open-ended questions.

Results: At the time of the survey, a total of 1458 patients had used AHR at least once. Mean age of AHR users was 47.5 years, 66.3% (967/1458) were male, 70.9% (1033/1458) were white, and 80.8% (1178/1458) had private insurance. Response rate was 26.8% (391/1458); 46.3% (180/389) were very satisfied with AHR, and 67.7% (264/390) described it as very easy to use. The most commonly utilized features were lab results (324/391, 82.9%), clinical vitals (264/391, 67.5%), and medications (253/391, 64.7%). No patient characteristics were associated with reporting high satisfaction or ease of use. The most common reason for using AHR was convenience/ease of use, and 58.2% (160/275) of users reported allowing no other apps to access their health information, citing privacy as one consideration.

Conclusions: Early adopters of AHR were demographically white, male, and privately insured. Convenience was an important facilitator, and users were selective in which apps they allowed to access their health information.

(J Med Internet Res 2022;24(1):e29367) doi:10.2196/29367

KEYWORDS

Apple Health Records; personal health record; electronic health record; patient satisfaction; early adopters; cross-sectional survey

Introduction

Despite near universal adoption of electronic health records (EHRs) in the United States, interoperability and data fragmentation remain significant challenges [1]. One proposed solution is the “personal electronic health record,” an internet-based set of tools that allow people to directly consolidate, access, and control their health information [1,2]. Development of a successful personal health record has been a policy objective since the creation of the National Coordinator for Health Information Technology in the Bush Administration [3].

Early attempts to develop personal health records by companies such as Google and Microsoft failed to achieve widespread adoption [3]. Lack of perceived usefulness, low trust, and a lack of meaningful integration of hospital data were among the

barriers limiting use [4]. In the last few years, a newer generation of mobile application–based personal health records has been developed. One example is Apple Health Records (AHR), a feature within the Apple Health app available on Apple iPhones that incorporates health care data from multiple health care sources and allows users to see integrated data simultaneously. In 2018, Apple partnered with 12 large health systems, including the University of Pennsylvania Health System (UPHS), to provide their patients with AHR. AHR has since expanded to more than 500 health systems [5]. AHR has significant potential for usage since iPhones are used by more than half the US population [6].

Yet, despite the number of health systems participating in the AHR pilot, little is known about the characteristics of AHR users and their experiences. To address this gap, we analyzed UPHS user characteristics and administered a survey to UPHS patients who had enabled the AHR feature.

Methods

Study Design

This study utilized a cross-sectional survey of AHR users at a large academic medical center. Survey results were linked to EHR data on demographics and health care utilization. Content analysis of open-response questions was performed.

AHR allows users to aggregate information from multiple health systems into a single interface within the iPhone health app. It gathers different forms of health information, including medications, vital signs, lab results, and procedures, integrating information from multiple sources into a single category for each information type. AHR thereby permits users to access organized information pooled from individual EHR systems on a single personal device. Users receive notifications when their health record is updated by a participating health system. Since this survey was administered, AHR has added features to perform longitudinal analysis of pooled data and share results with health care providers [7].

Human Subjects Research Review

This study was deemed excluded from human subjects review as a quality improvement project by the University of Pennsylvania institutional review board.

Participants

UPHS launched its pilot of AHR in January 2018. AHR is accessible through the Health application on an iPhone device. UPHS patients can add their UPHS data to the app using their UPHS patient portal credentials for authentication. Patients without an active portal account were directed to a website to create one. Participants included all UPHS users of AHR in the first 10 months of the pilot, defined as anyone who enabled AHR to download information to their mobile device.

Survey Instrument and Recruitment

A survey was developed by the study authors (see Multimedia Appendix 1). The survey was not based on a specific prior survey instrument but was developed to fit the intended purpose of capturing information from AHR users. The survey was designed using principles of survey design for Likert scale and open-response surveys and utilized standard categories for Likert scales. In addition, the survey was iteratively refined, initially through feedback from UPHS staff members, followed by pilot administration to a subsample of AHR users in September 2018. The final survey included 11 Likert scale, multiple selection, and open-ended questions. The survey ran from September 2018 to November 2018 and was administered using the Alchemer survey platform. Eloqua was used for email outreach and tracking. Users were sent an initial email with the survey link and up to 2 follow-up emails for nonresponse.

Demographic and Utilization Characteristics

Information was extracted on user age, sex, race, primary insurance, and health system utilization from the UPHS EHR system. Data were linked to survey responses, and the linked data were deidentified.

Analysis

Demographic and utilization characteristics were summarized for all users, and survey respondents were compared with nonrespondents. Differences between respondents and nonrespondents were compared using *t* tests for normal continuous variables, Wilcoxon rank-sum tests for nonnormal continuous variables, and chi-square tests for categorical variables. Patient portal users were compared with AHR users by identifying standardized differences greater than 0.2 [8]. This approach was used because the large sample size rendered even small differences statistically significant. Multivariable logistic regression was used to analyze the association of respondent characteristics with self-reported high satisfaction and ease of use. This analysis was considered exploratory, and model covariates were chosen based on demographic and utilization variables used to describe AHR and patient portal users. A multivariable analysis was performed in which age, hospitalizations, office visits, and active medications were included as continuous variables, sex was included as a binary variable, and race and primary insurance were included as categorical variables with the categories listed in Table 1. A *P* value <.05 was considered statistically significant. This threshold was set prior to all analyses. Satisfaction was dichotomized as very satisfied or not, and ease of use was categorized as reporting that AHR was very easy to use or not.

Open-ended responses were coded using content analysis. Two study investigators (JR and NP) separately coded all responses. Discrepancies were reviewed and resolved by mutual agreement.

Stata 14.2 was used for all data analyses.

Table 1. Demographic and utilization characteristics of users of Apple Health Records (AHR).

Characteristics	All patient portal users (N=535,422), n (%)	All AHR users (n=1458), n (%)	Survey responders[a] (n=373), n (%)	Survey nonresponders (n=1085), n (%)	*P* value[b] (n=1085)
Age (years), mean (SD)	50.0 (17.2)	47.5 (14.9)	50.0 (15.1)	46.6 (14.8)	<.001
Male, n (%)	195,223 (36.5)	967 (66.3)	289 (77.5)	678 (62.5)	<.001
Race, n (%)					
White	382,552 (71.4)	1033 (70.9)	296 (79.4)	737 (67.9)	<.001
Black	87,567 (16.4)	210 (14.4)	35 (9.4)	175 (16.1)	
Asian	22,915 (4.3)	90 (6.2)	16 (4.3)	74 (6.8)	
Other	21,791 (4.1)	78 (5.3)	19 (5.1)	59 (5.4)	
Unknown	20,597 (3.8)	47 (3.2)	7 (1.9)	40 (3.7)	
Primary insurance, n (%)					
Medicare	49,946 (9.3)	212 (14.5)	79 (21.2)	133 (12.3)	<.001
Medicaid	25,958 (4.9)	47 (3.2)	13 (3.5)	34 (3.1)	
None	7919 (1.5)	21 (1.4)	1 (0.3)	20 (1.8)	
Private	451,599 (84.3)	1178 (80.8)	280 (75.1)	898 (82.8)	
Hospitalizations[c], median (IQR)	0 (0-1)	0 (0-2)	0 (0-2)	0 (0-2)	.053
Office visits[c], median (IQR)	1 (0-3)	2 (1-5)	2 (1-5)	2 (1-5)	.63
Active prescriptions[c], median (IQR)	2 (0-6)	4 (1-10)	4 (0-11)	4 (1-10)	.82

[a]The number of respondents differs in Tables 1 and 2 because there were 18 survey responses with email addresses that could not be matched to a unique University of Pennsylvania Health System (UPHS) electronic health record. Reasons for nonmatching included patients who changed their UPHS email address and email addresses shared between multiple AHR users (eg, in the same household).

[b]Respondents compared with nonrespondents.

[c]Over 12 months prior to survey administration.

Results

Survey Responses and Respondent Characteristics

There was a total of 1458 UPHS AHR users at the time of survey administration. Mean user age was 47.5 (SD 14.9) years, 66.3% (967/1458) were male, and 70.9% (1033/1458) were White (Table 1). Median number of hospitalizations was 0 (IQR 0-2). Of 1458 users, 68.8% (1003/1458) opened the email (unique open rate), 29.1% (424/1458) accessed the survey (click-through rate), and 26.8% (391/1458) submitted a response. Survey respondents were older, more likely to be male, more likely to be White, and more likely to have private insurance than survey nonrespondents. There were no statistically significant differences in hospitalizations, office visits, or prescriptions. Although AHR users were majority male, in contrast, only a minority of patient portal users were male (195,223/535,422, 36.5%; standardized difference 0.63). In addition, AHR users had higher utilization than all patient portal users over 12 months, including more hospitalizations (standardized difference 0.21), more office visits (standardized difference 0.25), and more active prescriptions (standardized difference 0.21).

Of the 391 participants, 180 (46.3%) were very satisfied with AHR, whereas 10 (2.6%) were very dissatisfied (Table 2). Although 264 (67.7%) of the 391 participants described AHR as very easy to use, 6 (1.5%) called it very difficult. On a 1 (low) to 9 (high) scale, respondents reported a mean 7.7 (1.9) likelihood to recommend to a friend. The most commonly used features were lab results (324/391, 82.9%), clinical vitals (264/391, 67.5%), and medications (253/391, 64.7%). Respondents most often shared information with family (191/391, 48.8%), followed by not sharing with anyone (138/391, 35.3%) and sharing with a physician (137/391, 35.0%). The most frequent means of finding out about AHR was through news media (142/391, 36.3%).

Table 2. User survey responses.

Response	Results
Overall satisfaction, n (%)	
Very satisfied	180 (46.3)
Satisfied	152 (39.1)
Neutral	39 (10.0)
Dissatisfied	8 (2.1)
Very dissatisfied	10 (2.6)
How easy to use, n (%)	
Very easy	264 (67.7)
Somewhat easy	99 (25.4)
Neutral	18 (4.6)
Difficult	3 (0.8)
Very difficult	6 (1.5)
How likely to recommend to a friend (1-9), mean (SD)	7.7 (1.9)
Features used, n (%)	
Lab results	324 (82.9)
Clinical vitals	264 (67.5)
Medications	253 (64.7)
Conditions	183 (46.8)
Procedures	173 (44.2)
Immunizations	160 (40.9)
Allergies	141 (36.1)
All records	14 (3.6)
Other	13 (3.3)
Discussed information in AHR[a], n (%)	
Family	191 (48.8)
Did not share with anyone[b]	138 (35.3)
Physician	137 (35.0)
Friend	76 (19.4)
Other members of care team	56 (14.3)
Pharmacists	19 (4.9)
Other person	6 (1.5)
How found out about AHR, n (%)	
News article	142 (36.3)
Email announcement	75 (19.2)
Friend	20 (5.1)
Family	14 (3.6)
Physician	10 (2.6)
Other member of care team	6 (1.5)

[a]AHR: Apple Health Records.

[b]Did not share results (ie, no one).

In adjusted models, no demographic or utilization characteristics were significantly associated with reported satisfaction with AHR or ease of use (Table 3).

Table 3. Association of user characteristics with ratings of Apple Health Records.

Characteristics	Very satisfied, OR[a] (95% CI)	*P* value	Very easy to use, OR (95% CI)	*P* value
Age	1.00 (0.99-1.02)	.79	1.00 (0.98-1.02)	.83
Male	0.89 (0.53-1.49)	.65	0.83 (0.47-1.46)	.52
Race	1.05 (0.84-1.31)	.69	1.03 (0.81-1.32)	.80
Primary insurance	1.13 (0.87-1.46)	.37	1.19 (0.91-1.56)	.19
Hospitalizations	0.87 (0.65-1.15)	.32	0.88 (0.65-1.19)	.40
Office visits	1.00 (0.94-1.08)	.90	0.99 (0.92-1.07)	.79
Active prescriptions	1.00 (0.99-1.02)	.66	1.01 (0.99-1.02)	.45

[a]OR: odds ratio.

Responses to Open-Ended Questions

The most common reason that respondents gave for using AHR was convenience/ease of use (138/305, 45.2%), followed by having all information in one place (54/305, 17.7%; Table 4). One respondent noted that "[i]t was very easy and I don't have to log in to view them." Asked what other information they wanted to be included, respondents indicated, at near equal rates, no other information and all medical information. For example, one respondent said "[i]'d like to have EVERYTHING that Penn Medicine has in my patient record accessible through Apple Health Records." The most frequent request for change to AHR was to provide better labeling and displays.

When asked what other apps respondents allowed to access their health information, the most common answer (160/275, 58.2%) was "none." Privacy was cited as one reason. One respondent explained: "As you know, privacy is a major concern with health records. Apple has a reputation for being very serious about customer privacy. I would never grant access to health records to companies like Google or Facebook, whose business model rests on selling privacy."

Table 4. Coded responses to open-response questions.

Questions and responses	Responses, n (%)[a]	Example quotations
Approximately how much time have you spent using Apple Health Records?[b] (n=222)		
Not much	32 (14.4)	-[c]
<1 hour	134 (60.4)	"About a half-hour or so reviewing records."
1 to 2 hours	28 (12.6)	-
2 to 3 hours	10 (4.5)	-
≥3 hours	18 (8.1)	-
Why did you choose to access your Penn health information via Apple Health Records? (n=305)		
Convenience/ease of use	138 (45.2)	"It was very easy and I don't have to log in to view them"
Information in one place	54 (17.7)	"I love having all of my information in one place. I don't want to have an app for each provider. I am confident that Apple is keeping measures to keep my information secure"
Experimental/curiosity	47 (15.4)	-
Apple production	42 (13.8)	-
Portability	13 (4.3)	-
Access specific information	11 (3.6)	-
What other health information would you like to have through your Apple Health Records? (n=181)		
No other information	50 (27.6)	-
Everything	49 (27.1)	"I'd like to have EVERYTHING that Penn Medicine has in my patient record accessible through Apple Health Records."
Appointments	26 (14.4)	-
Radiology	26 (14.4)	-
Clinical notes	13 (7.2)	-
Additional lab data	10 (5.5)	-
Communication with care team	7 (3.9)	-
If you could change one thing about Apple Health Records, what would it be? (n=170)		
No changes	63 (37.1)	-
Better display, labeling, and analysis	44 (25.9)	"Better visualization of test results, etc. changing over time. Graphs would be good."
Add providers and health systems	23 (13.5)	-
Access to actual reports, imaging, and documentation	17 (10)	-
Integration with information from other sources	16 (9.4)	-
Change units	7 (4.1)	-
What other apps have you allowed to access your health record data? (n=275)		
None	160 (58.2)	"None. As you know, privacy is a major concern with health records. Apple has a reputation for being very serious about customer privacy. I would never grant access to health records to companies like Google or Facebook, whose business model rests on selling privacy."
Fitness trackers	69 (25.1)	-
Patient portal	23 (8.4)	-
Other health systems patient portals	13 (4.7)	-
Other Apple products	10 (3.6)	-

[a]Totals differ by question due to nonresponse, and numbers for each question with codable responses are shown.

[b]Time period refers to overall usage as an Apple Health Record user.

[C]Example not listed.

Discussion

Principal Findings

In this survey of patients using AHR at an early health system adopter, most respondents reported high satisfaction and ease of use, and none of the tested characteristics was associated with differences in ratings. Convenience/ease of use was a frequently cited rationale for use, and many respondents requested incorporation of all health information into AHR. This study has several notable findings for future use of mobile application–based personal health records such as AHR.

First, personal health records may be used by specific patient demographics. Most patients were White and privately insured. In comparison with all patient portal users, AHR users were disproportionately male. Disparities are a concern with mobile application–based personal health records, because smartphone users are younger and wealthier than the US population [9,10]. Older patients have lower digital health literacy than younger patients [11], although such disparities are decreasing over time [12]. Yet, demographics also were not associated with differences in AHR ratings. In addition, although AHR users had higher health care utilization than patient portal users, utilization remained low, with a low number of office visits and predominantly no hospitalizations in the year prior to the study, suggesting that the app was not reaching patients with high health system utilization who may have the most significant need to share information because of frequent hospitalizations or complex care. The gender disparities in use are surprising given the female-predominant overall patient portal population and require further investigation to understand how barriers and facilitators of use disproportionally affect female patients.

Second, convenience was an important facilitator of use, and most AHR users were selective about which apps accessed their health information. For personal health records to achieve widespread use, they will need to overcome several barriers. Platforms will need to be convenient and provide the privacy protections for patients to feel comfortable with storage of substantial health information.

Indeed, privacy considerations and trust in the Apple brand to safeguard information emerged as important factors, a finding consistent with past studies of personal health records [13]. Safeguarding of personal information will be crucial to personal health record success, as security lapses leading to stories of compromised personal information may raise privacy concerns among would-be users.

Comparison With Prior Work

These findings advance our understanding of personal health records. Prior work has examined utilization and factors affecting adoption of national personal health records in countries such as England and Portugal [14,15]. Other work sought to understand the barriers to adoption of earlier personal health records such as Google Health [4]. In addition, commentators have noted the potential advantages of mobile application–based personal health records offered by private vendors, including AHR [1,16]. This study offers new insight into the characteristics and perspectives of AHR users at one of the first health systems to participate in the AHR pilot.

Empowering patients with their health records has been a goal for the Office of the National Coordinator (ONC) for Health Information Technology. In 2016, Congress passed the 21st Century Cures Act to drive electronic access, exchange, and use of health information. The ONC Cures Act Final Rule implemented the interoperability provisions of the Cures Act to promote patient control over their own health information [17].

The rule requires the health care industry to adopt standardized application programming interfaces (APIs) to allow patients to securely access structured electronic health information using smartphone applications. These changes will make it easier for third-party apps to access and integrate health data, including AHR, which uses these same standards for data exchange. AHR is utilizing these standards to add new data-sharing features, including bidirectional sharing that will allow users to directly share AHR data with certain AHR systems [18]. Understanding the barriers and facilitators of using these tools is crucial to guide future development in a direction that will meet the needs of patients and potentially address health disparities that could result from wider adoption. It will also be necessary for information policy development in this area.

Limitations

This study has several limitations. It reported results from a single-center experience with one platform. Results may not generalize to other health systems or platforms. That said, this early experience with AHR adds to our knowledge of an important new exemplar of personal health records. This study also found some differences between respondents and nonrespondents, and respondents may not be representative of the overall user base. Further efforts will be needed with larger-scale surveys and qualitative studies. Third, this study provided only information from self-reports. Finally, this study examined only patient perspectives and not how AHR use might affect clinical outcomes.

In summary, we reported characteristics and perspectives of patient users at an early adoption of AHR. Our experiences offer several lessons for future use and study of personal health records.

Acknowledgments

We thank William Hanson and Christine VanZandbergen for their guidance and support. We thank Sarah LaMar for data support.

Authors' Contributions

NP developed the initial idea for the study. All authors participated in study development and design. JR, RW, and NP developed the survey instrument. GT and RW collected and organized the data for analysis. JR wrote the initial draft of the manuscript. All authors revised the manuscript for critical content.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental survey.
[DOCX File , 16 KB - jmir_v24i1e29367_app1.docx]

References

1. Dameff C, Clay B, Longhurst CA. Personal health records: more promising in the smartphone era? JAMA 2019 Jan 29;321(4):339-340. [doi: 10.1001/jama.2018.20434] [Medline: 30633300]
2. Connecting for health: a public-private collaborative. Markle Foundation. 2003 Jul 01. URL: https://www.markle.org/sites/default/files/statementofpurpose.pdf [accessed 2021-12-18]
3. Braunstein ML. Health care in the age of interoperability part 5: the personal health record. IEEE Pulse 2019 May;10(3):19-23. [doi: 10.1109/mpuls.2019.2911804]
4. Spil T, Klein R. Personal health records success: why Google Health failed and what does that mean for Microsoft HealthVault? 2014 Presented at: 47th Hawaii International Conference on System Sciences; January 6-9, 2014; Waikola, HI. [doi: 10.1109/hicss.2014.353]
5. Institutions that support health records on iPhone and iPod touch. Apple. URL: https://support.apple.com/en-us/HT208647 [accessed 2021-12-18]
6. Number of Apple iPhone users in the U.S. 2012-2022. Statista. 2021 Mar 01. URL: https://www.statista.com/statistics/232790/forecast-of-apple-users-in-the-us/ [accessed 2021-12-18]
7. Muoio D. Apple unveils automated health data alerts, in-EHR record sharing, fall risk analysis and more at WWDC 2021. Fierce Healthcare. 2021 Jun 07. URL: https://www.fiercehealthcare.com/digital-health/apple-unveils-automated-health-data-alerts-ehr-record-sharing-fall-risk-analysis-and [accessed 2021-12-18]
8. Yang D, Dalton JE. A unified approach to measuring the effect size between two groups using SAS. SAS Global Forum. 2012. URL: https://www.lerner.ccf.org/qhs/software/lib/stddiff.pdf [accessed 2021-12-18]
9. Vogels EA. Digital divide persists even as Americans with lower incomes make gains in tech adoption. Pew Research Center. 2021 Jun 22. URL: https://www.pewresearch.org/fact-tank/2019/05/07/digital-divide-persists-even-as-lower-income-americans-make-gains-in-tech-adoption/ [accessed 2021-12-18]
10. Mobile Fact Sheet. Pew Research Center. 2021 Apr 07. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-12-18]
11. Levine DM, Lipsitz SR, Linder JA. Trends in seniors' use of digital health technology in the United States, 2011-2014. JAMA 2016 Aug 02;316(5):538-540. [doi: 10.1001/jama.2016.9124] [Medline: 27483069]
12. Tarver WL, Menser T, Hesse BW, Johnson TJ, Beckjord E, Ford EW, et al. Growth dynamics of patient-provider internet communication: trend analysis using the Health Information National Trends Survey (2003 to 2013). J Med Internet Res 2018 Mar 29;20(3):e109 [FREE Full text] [doi: 10.2196/jmir.7851] [Medline: 29599107]
13. Roehrs A, da Costa CA, Righi RDR, de Oliveira KSF. Personal health records: a systematic literature review. J Med Internet Res 2017 Jan 06;19(1):e13 [FREE Full text] [doi: 10.2196/jmir.5876] [Medline: 28062391]
14. Laranjo L, Rodolfo I, Pereira AM, de Sá AB. Characteristics of innovators adopting a national personal health record in Portugal: cross-sectional study. JMIR Med Inform 2017 Oct 11;5(4):e37 [FREE Full text] [doi: 10.2196/medinform.7887] [Medline: 29021125]
15. Abd-Alrazaq A, Bewick BM, Farragher T, Gardner P. Factors affecting patients' use of electronic personal health records in England: cross-sectional study. J Med Internet Res 2019 Jul 31;21(7):e12373 [FREE Full text] [doi: 10.2196/12373] [Medline: 31368442]
16. Neinstein A, Thao C, Savage M, Adler-Milstein J. Deploying patient-facing application programming interfaces: thematic analysis of health system experiences. J Med Internet Res 2020 Apr 03;22(4):e16813 [FREE Full text] [doi: 10.2196/16813] [Medline: 31983680]
17. 21st Century Cures Act: Interoperability, Information Blocking, and the ONC Health IT Certification Program. Federal Register. 2020 May 01. URL: https://www.federalregister.gov/documents/2020/05/01/2020-07419/21st-century-cures-act-interoperability-information-blocking-and-the-onc-health-it-certification [accessed 2021-12-18]
18. Drees J. Cerner partners with Apple on new health records sharing initiative: 5 details. Becker's Health IT. 2021 Jun 09. URL: https://www.beckershospitalreview.com/ehrs/cerner-partners-with-apple-on-new-health-records-sharing-initiative-5-details.html [accessed 2021-12-18]

Abbreviations

AHR: Apple Health Records
API: application programming interface
EHR: electronic health record
ONC: Office of the National Coordinator
UPHS: University of Pennsylvania Health System

Edited by R Kukafka; submitted 10.04.21; peer-reviewed by A Pereira, M Savage; comments to author 26.05.21; revised version received 28.08.21; accepted 21.11.21; published 25.01.22.

Please cite as:
Rolnick J, Ward R, Tait G, Patel N
Early Adopters of Apple Health Records at a Large Academic Medical Center: Cross-sectional Survey of Users
J Med Internet Res 2022;24(1):e29367
URL: https://www.jmir.org/2022/1/e29367
doi:10.2196/29367
PMID:35076397

Original Paper

Integrating Quality of Life in the Care Pathway of Cancer Patients Undergoing Immunotherapy Treatment: Descriptive, Cross-sectional Survey of an Online Patient Community's Experiences and Expectations

Ophélie Wilczynski[1], MSci; Anthony Boisbouvier[1], MSc; Lise Radoszycki[1], MSci; François-Emery Cotté[2], PharmD, PhD; Anne-Françoise Gaudin[2], PharmD; Hervé Lemasson[2], MSc

[1]Carenity, Paris, France

[2]Bristol-Myers Squibb France, Rueil-Malmaison, France

Corresponding Author:
Hervé Lemasson, MSc
Bristol-Myers Squibb France
3 Rue Joseph Monier
Rueil-Malmaison, 92506
France
Phone: 33 1 58 83 60 00
Email: Herve.Lemasson@bms.com

Abstract

Background: New cancer treatments, such as immune checkpoint inhibitors (ICIs), can improve survival and health-related quality of life (HRQoL) in patients with cancer. Although long-term monitoring of HRQoL has been shown to improve survival, integration of HRQoL into everyday practice remains poorly documented.

Objective: This study describes experiences and expectations of patients treated with ICIs regarding a discussion of HRQoL with health care professionals (HCPs) in cancer management.

Methods: This cross-sectional study was conducted in an online patient community (Carenity) in France. Patients treated with ICIs for cancer, included between September 2018 and January 2019, completed a questionnaire to assess the involvement of HCP in a discussion of HRQoL and when and what was discussed.

Results: Of 82 patients included (mean age: 56.9 years, 95% CI 54.2-59.6; 46 [56%] male; 34 [41%] with lung cancer), 62 (76%) reported discussing HRQoL at least once with HCPs, mainly general practitioners (54/82, 66%), oncologists (53/82, 65%), and hospital nurses (50/82, 61%). Around half (45/82, 55%) of the patients were satisfied with these discussions. Discussions with the oncologist were at the patient's initiative (34/53, 64%). Discussions occurred primarily during follow-up visits (40/62, 65%), when adverse events occurred (30/62, 48%), and at treatment initiation (27/62, 32%). The most discussed dimensions were symptoms (48/62, 77%) and physical well-being (43/62, 69%). With respect to expectations, 54/82 (66%) patients considered oncologists as the most important HCPs for discussing HRQoL. These discussions were desirable throughout the care pathway, particularly at diagnosis (63/82, 77%) and when treatment was initiated (75/82, 92%) or changed (68/82, 83%). All HRQoL dimensions were considered important to discuss.

Conclusions: With only around half of the patients satisfied with HRQoL discussions, impactful HRQoL integration in clinical practice is critical. According to patients, this integration should involve mainly oncologists and general practitioners, should happen at every step of the care pathway, and should be extended to dimensions that are currently rarely addressed.

(J Med Internet Res 2022;24(1):e25792) doi:10.2196/25792

KEYWORDS

cancer; quality of life; immunotherapy; patient community; patient satisfaction

Introduction

Health-related quality of life (HRQoL) is a critical feature of the life of patients with cancer, and a number of instruments have been developed for evaluating this over the past 40 years. These include general HRQoL instruments that are not specific to cancer but can be used to compare HRQoL between cancer and other diseases, as well as instruments that are specific to cancer [1]. Such cancer-specific instruments include the European Organisation of Research and Treatment of Cancer Quality of Life Questionnaire (EORTC-QLQ) family of questionnaires [2] and the Functional Assessment of Cancer Therapy (FACT) family [3], of which individual versions have been designed for specific types of cancer. These disease-specific HRQoL measures are used systematically as outcome measures in clinical trials but may also be used to support discussions of HRQoL in the everyday care of patients with cancer.

In cancer, HRQoL is impacted by disease symptoms as well as by side effects and constraints associated with therapy. Moreover, HRQoL can change rapidly and sometimes unpredictably over the course of the disease. Preserving the HRQoL of patients with cancer is a major goal of clinicians and health authorities [4-6]. For this reason, building and maintaining an open dialogue between patients and health care professionals (HCPs) is essential in order to evaluate the patient's HRQoL adequately and to promptly address any issues that may arise. Systematic monitoring of the patient's perceptions of HRQoL has also been shown to be of benefit in terms of symptom management [7], satisfaction with their care [7], a greater use of supportive care [7], improvement in clinician-patient communication [8,9], and improved overall survival [10-13], since it allows, among other potential advantages, timely adaptation of treatment in the case of symptom progression or emergence of treatment side effects.

The introduction of immune checkpoint inhibitors (ICIs) over the past decade has represented a major advance in the treatment of many types of cancer, allowing sustained recovery and, for some tumors, potentially elimination of disease in a significant proportion of patients [14,15]. By providing patients with a survival benefit [16] and a better tolerance profile compared to traditional chemotherapy [17], treatment with ICIs has become an attractive therapeutic alternative for many types of cancer. In terms of HRQoL, the experience of cancer patients treated with immunotherapy may differ from that of patients receiving standard chemotherapy. Treatment with ICIs may be associated with a different profile of response compared with standard chemotherapy, due to longer periods of disease stability and the lower incidence of side effects that have an impact on the quality of life (QoL) [17]. Several studies have investigated HRQoL in patients treated with ICIs [18-20] and have shown maintenance of HRQoL over long periods, and even improvement in HRQoL compared to standard chemotherapy in certain patients [18,19,21,22].

Most QoL research with ICIs has been conducted in the context of clinical trials, although some observational studies on long-term survivors who are clinically stable have been reported [23,24]. Potentially deleterious effects of ICI-specific adverse events on HRQoL [25,26] and potentially beneficial effects on social functioning and role integration [27-29] are aspects that would deserve attention. In addition, development of a specific HRQoL measure for cancer patients using ICIs with stable disease could be useful [30]. From an operational and a care perspective, the changes in the treatment paradigm associated with the introduction of ICIs indicate the utility of monitoring HRQoL over the long term in everyday practice. This could provide benefits in optimizing functional outcomes in a timely manner, as well as in contributing to treatment decisions. However, little information is available on how HRQoL is considered by physicians treating cancer patients with ICIs in routine clinical practice.

The objectives of this study were to describe experiences of patients treated with ICIs and their expectations with respect to how the importance of HRQoL in cancer management is considered by HCPs. This includes the description of practices of HCPs in appraising HRQoL with patients currently or previously treated with ICIs, the evaluation of patient satisfaction with their dialogue about HRQoL with their HCPs, and the identification of patient expectations with respect to discussing HRQoL.

Methods

This study was a descriptive, cross-sectional web-based survey of cancer patients (or their relatives) treated with ICIs who were members of the Carenity cancer community and resident in France. Participation was voluntary. Participants were recruited over 4 months from September 10, 2018, to January 7, 2019.

Study Population

The study population included participants from the Carenity cancer community. Carenity is an online patient community for people with chronic conditions [31,32]. Patients and caregivers can share their experiences in more than 1200 disease-specific communities, exchange information on the disease and request advice and information. They can also participate in online surveys concerning various aspects of disease perceptions on a voluntary basis and after giving explicit consent. Currently, the cancer patient community on Carenity in France has around 9547 members, of whom 5871 (61.5%) are patients. All Carenity cancer community members were invited to participate in this study.

Participants could either be patients themselves or a relative (or friend) who was prepared to complete the study questionnaire on their own or with the patient. Relatives were asked to complete the questionnaire from the patient's point of view. In the rest of the manuscript, the data presented represent the characteristics and opinions of the patients, regardless of whether it was the patients themselves or a relative who completed the questionnaire. Participants were eligible for the study if they or their relative were currently or previously treated with an ICI (atezolizumab, durvalumab, nivolumab, pembrolizumab, or ipilimumab).

As members of the Carenity platform, patients or relatives participating in the study provided explicit informed consent to

the collection, handling, and keeping of their personal and health data. They were also provided with specific information about the goals and procedures of the study, as well as about the notion of HRQoL, and asked to agree to participate before receiving the study questionnaire. Participants received no incentives to participate in the study, and participation had no impact on their future involvement as Carenity platform members.

Study Questionnaire

The questionnaire was developed specifically for this study. The HRQoL domains explored are based on constructs in 2 existing validated cancer-specific HRQoL questionnaires (QLQ-C30 [33] and FACT-G [3]). The questionnaire was subsequently tested for clarity and relevance by 2 representatives of the Carenity cancer community.

The study questionnaire started with a set of screening questions to identify the participant as a patient or as a relative and to ensure that the patient had a diagnosis of cancer and was being treated (or had been treated previously) with an ICI. If this was not the case, the participant left the study at this point. Otherwise, they proceeded to the core questionnaire, which took, on average, around 15 minutes to complete.

The core questionnaire consisted of 29 questions for all participants, as well as 3 additional ones to be completed only by relatives answering on behalf of a patient, which were divided into 3 sets, relating to general information, experiences with discussion of HRQoL, and expectations for discussing HRQoL with HCPs. The themes and attributes evaluated during the study are listed by theme in Table 1. The first set of 14 general questions collected data on patient demographics, cancer history, recent treatment (12 months), HCPs consulted, and treatment location. The second set of questions started with an open-ended question asking patients to sum up in 3 words or phrases the aspects of their HRQoL that were most impacted by cancer and its treatments. Participants were then asked whether they had ever discussed QoL with an HCP and when. The period of time covered was not restricted to the period of treatment by ICIs but related to the entire period since the diagnosis of cancer was given. Only patients for whom this was the case completed the other questions in this set. in total, 10 questions collected information about the dialogue between the patient and the medical care team, covering the type of HCP involved, when HRQoL was discussed, the aspects of HRQoL discussed, and satisfaction with the discussions. Finally, all participants, whether or not they had discussed HRQoL, completed the last set of 6 questions about expectations for a dialogue with an HCP about QoL, which covered an identical set of concepts as those explored in the previous set of questions on experiences.

Table 1. Information collected during the study.

Attribute studied	Question	Response modality/data analysis	Data presentation
Perceptions of the impact of cancer on QoL[a]			
Impact of cancer on QoL	Can you cite 3 words or expressions that you think best express the aspects of QoL that are impacted by your cancer?	• Open question • Replies grouped by theme	• Number and % of patients citing each theme
Discussion of QoL with HCPs[b]			
Importance of discussing QoL	Do you think that discussing QoL with HCPs is . . . (list)?	• Checklist of 5 levels of importance • Single response only	• Number and % of patients citing each importance level
Experience of discussing QoL	On what occasion(s) did you discuss QoL with the HCP who looks after you?	• Checklist including "Never" • Multiple responses possible	• Number and % of patients citing each occasion • Number of different HCPs identified
Desire to discuss QoL	You replied that you have never discussed QoL with an HCP. Would you have liked an opportunity to do so?	• Yes/No/Don't know	• Number and % of patients replying yes
Satisfaction with discussions of QoL	Were you satisfied with the way that QoL has been brought up by different HCPs?	• Checklist • Single response only	• Number and % of patients citing each response
Reasons for satisfaction or dissatisfaction	What was the reason that you were satisfied or dissatisfied?	• Open question • Replies grouped by theme	• Number and % of patients citing each theme
Opportunity to express yourself	Do you feel that you were able to express yourself about the impact of cancer or cancer treatments on your QoL?	• Checklist of 5 response modalities • Single response only	• Number and % of patients citing each response
HCPs involved in HRQoL[c] discussions			
Types of HCP discussing QoL	When you consult 1 of the following types of HCP, do you discuss QoL with them?	• Checklist of different HCPs with 5 response modalities for each • Including "Never/ I don't consult this HCP" • Single response only	• Number and % of patients responding often, occasionally, or rarely for each HCP specialty
Who initiates the discussion?	When you discuss QoL with your oncologist or radiotherapist, who usually initiates the conversation?	• Checklist of 5 response modalities • Single response only	• Number and % of patients citing each response
Importance of different HCPs	Which HCPs do you think are the most important for talking about QoL?	• Checklist of different HCPs • Multiple responses possible	• Number and % of patients citing each type of HCP
Other contexts where QoL is discussed	Have you ever discussed your QoL in another context (discussion group, therapeutic education program, etc)?	• Checklist of 5 contexts • Single response only	• Number and % of patients citing each context
Opportunities for discussing QoL			
Occasions when QoL had been discussed	On what occasion(s) did you discuss QoL with the HCP who looks after you?	• Checklist • Multiple responses possible	• Number and % of patients citing each occasion
Relative importance of different occasions for discussing QoL	Which occasions do you think are particularly important for discussing QoL with HCPs?	• Checklist of different HCPs with 5 response modalities for each • Single response only	• Number and % of patients citing each occasion
Dimensions of QoL discussed			
Subjects discussed	When you discuss QoL, what are the subjects that you usually discuss?	• Checklist • Multiple responses possible	• Number and % of patients citing each subject • Number of different subjects identified

Attribute studied	Question	Response modality/data analysis	Data presentation
Relative importance of discussing different subjects	How much importance do you attach to discussing the following subjects with an HCP?	• 10 cm visual analog scale for each of the 9 subjects	• Mean score with standard deviation
Measures for improving discussions of QoL			
Ways to improve paying attention to QoL	How could the medical team involved in your care pay more attention to your QoL?	• Open question • Replies grouped by theme	• Number of citations for each theme
Specific measures	In your opinion, which are the 3 measures that would be most useful to improve discussions of your QoL?	• Checklist of 11 measures • 3 responses possible	• Number and % of patients citing each measure

[a]QoL: quality of life.

[b]HCP: health care professional.

[c]HRQoL: health-related quality of life.

All questions in this web-based survey were mandatory to access the next question, except for 4 open ones. Skip patterns were used, when appropriate. Most of the questions were single or multiple choice, to which participants responded by ticking boxes. Two questions were in the form of Likert scales, and another asked patients to rate the importance of 9 HRQoL dimensions on a visual analog scale.

Data Analysis

The data analysis was purely descriptive as no prespecified hypotheses were tested. Responses to multiple-choice questions and Likert scales are presented as frequency counts and percentages with their 95% CIs.

Ethics

The study was conducted in accordance with good epidemiological practice. As the aim of the study was to determine patient satisfaction, the survey is considered a patient satisfaction survey and does not fall within the scope of French legislation on medical research. For this reason, submission to an ethical committee was not required.

Results

Patient Population

A total of 82 questionnaires were fully completed, of which 56 (68%) were completed by the patients and 26 (32%) by a friend or relative. In the latter case, 16 (61%) questionnaires were completed in the presence of the patient. The characteristics of the patients are presented in Table 2. Overall, 46 of 82 (56%) patients were men, and the most frequent cancer types were lung cancer, lymphoma, and skin cancer, which accounted between them for 58 (71%) cases. The remaining cancer types accounted for ≤5 (6%) patients each. The mean age was 56.9 years (95% CI 54.2-59.6), and this was similar across the principal cancer types (58 years for lung cancer and lymphoma and 52 years for skin cancer). The diagnosis of cancer had been made within the previous 5 years for two-thirds of patients. Overall, 62 of 82 (76%) patients had discussed their HRQoL with an HCP, and only these patients completed the set of questions about their experience. Information about 1 patient who had died was provided by a relative.

Table 2. Characteristics of study patients (N=82).

Characteristic	n (%)
Age (years)	
18-30	2 (2%)
31-40	7 (9%)
41-50	12 (15%)
51-60	29 (35%)
61-70	22 (27%)
>70	10 (12%)
Gender	
Men	46 (56%)
Women	36 (44%)
Primary cancer localization	
Lung	34 (41%)
Lymphoma	12 (15%)
Skin	12 (15%)
Kidney	5 (6%)
Prostate	3 (4%)
Ovarian	3 (4%)
Leukemia	3 (4%)
Other[a]	10 (12%)
Time since cancer diagnosis	
0-5 years	65 (79%)
6-10 years	10 (12%)
>10 years	6 (7%)
Do not know	1 (1%)
Place of treatment in previous 12 months[b]	
University hospital	28 (34%)
Local hospital	27 (33%)
Private clinic	25 (30%)
Specialist cancer center	14 (17%)
Community medical center	4 (5%)
Not treated in previous 12 months	1 (1%)

[a]Head and neck, multiple myeloma, and bladder cancer: 2 cases each; colon, liver, cervical, and bladder/prostate cancer: 1 case each.

[b]Multiple responses were possible.

Perceptions of the Impact of Cancer on QoL

For the aspects of QoL that were most impacted by cancer and its treatment, the theme that was most frequently cited was physical well-being, cited by 52 of 82 (63%) patients. In addition, impact on activities of daily living and emotional well-being were also frequently mentioned, by 25 of 82 (30%) patients each. The most frequent responses cited in the physical-well-being theme were fatigue (26 citations), difficulty getting about (13 citations), and pain (12 citations). The most frequent responses cited in the activities-of-daily-living theme were shopping (10 citations), washing (6 citations), and do-it-yourself/gardening (5 citations). The most frequent responses cited in the emotional-well-being theme were mood (18 citations), stress/anxiety (8 citations), and solitude (4 citations). A full listing of the themes evoked is provided in Table 3.

Table 3. Themes of quality of life most impacted by cancer (N=82).

Theme	Number of citations, n	Number of patients citing theme, n (%, 95% CI)
Physical well-being		
Total	74	52 (63%, 53%-74%)
Fatigue	26	—[a]
Difficulty getting about	13	—
Pain	12	—
Difficulty sleeping	6	—
Difficulty breathing	4	—
Difficulty in the morning	4	—
Concentration	2	—
Weight gain	2	—
Incontinence/diarrhea	2	—
Loss of appetite	1	—
Sensitivity to changes in the weather	1	—
Falling ill more often	1	—
Activities of daily living		
Total	36	25 (30%, 21%-41%)
Shopping	10	—
Washing/dressing	6	—
Gardening/jobs in the house	5	—
Cleaning	4	—
Driving	4	—
Cooking	3	—
Daily activities	3	—
Keeping appointments	1	—
Emotional well-being		
Total	33	25 (30%, 21%-41%)
Daily morale	18	—
Stress/anxiety	8	—
Loneliness	4	—
Motivation	1	—
Fear of dying	1	—
Feeling helpless	1	—
Leisure activities		
Total	23	21 (26%, 16%-35%)
Sport/physical activity	11	—
Going walking	6	—
Leisure	3	—
Dancing	2	—
Traveling	1	—
Social and family life		
Total	21	20 (24%, 15%-34%)
Outings	6	—

Theme	Number of citations, n	Number of patients citing theme, n (%, 95% CI)
Family	6	—
Sex life	4	—
Seeing friends	3	—
The way people look at me	1	—
Conversation	1	—
Professional life		
Total	6	6 (7%, 2%-13%)
Others		
Total	9	9 (11%, 4%-18%)
Long-term planning	2	—
Autonomy	2	—
Wasting time	2	—
Finding a doctor	1	—
Not doing anything any more	1	—
Organization	1	—

[a]Not applicable.

Discussion of QoL with HCPs

Overall, 75 of 82 (91%) patients considered it important to discuss their HRQoL with an HCP, with 58 (71%) considering it very important and a further 16 (20%) considering it quite important. In addition, 62 of 82 patients (76%) patients had discussed their HRQoL with an HCP at least once. Of the 20 patients who had not done so, 9 (45%) would have liked to, 4 (20%) were not interested, and the remaining 7 (35%) did not know. In addition, 45 of 82 (55%) patients were always or often satisfied with the way in which their HRQoL had been discussed. The principal reasons for satisfaction were that the discussion had resulted in practical solutions being identified (26/45, 58%) and a good relationship with the HCP due to their human qualities (24/45, 53%). Of the 82 patients, 17 (21%) were, however, frequently dissatisfied with this discussion. Reasons for dissatisfaction were insufficient time available for discussing HRQoL (9/17, 53%), a lack of information and explanations provided by the HCP (5/17, 29%), and a lack of empathy on the part of the HCP (4/17, 24%). In addition, 29 of 82 (35%) patients considered that they had been listened to when discussing their HRQoL, whereas an identical number considered that they had not been sufficiently listened to or given the chance to express themselves.

HCPs Involved in HRQoL Discussions

Patients reported discussing HRQoL with a variety of different HCPs, with the majority reporting multiple points of contact. On average, patients reported consulting 6.7 (95% CI 6.2-7.2) different types of HCPs and discussing HRQoL with, on average, 5.8 (95% CI 5.3-6.3) of these. The most frequently cited HCPs were the general practitioner, the oncologist or radiologist, and the hospital nurse (Table 4). It should be noted that certain HCPs who are frequently consulted, such as community nurses and pharmacists, less frequently discuss HRQoL, whereas other HCPs generally do discuss this issue, even though they are less frequently consulted, such as psychiatrists or palliative care physicians. For 34 of the 53 (64%) patients discussing HRQoL with their oncologist or radiologist, the discussion was initiated by the patient rather than by the physician. When patients were asked with which sort of HCP it was important to discuss HRQoL, the oncologist or radiologist and the general practitioner were the 2 professions that were most often cited, followed by other specialist physicians, the psychiatrist or psychologist, and the hospital nurse (Table 4).

Table 4. Health care professionals discussing quality of life with patients (N=82).

Type of HCP[a]	Number of patients who consulted indicated HCP, n (%, 95% CI)	Number of patients who discussed HRQoL[b] with indicated HCP[c], n (%, 95% CI)	Number of patients who considered indicated HCP important for discussions of HRQoL, n (%, 95% CI)
General practitioner	57 (70%, 60%-80%)	54 (66%, 56%-76%)	44 (54%, 43%-64%)
Oncologist or radiologist	56 (68%, 58%-78%)	53 (65%, 54%-75%)	54 (66%, 56%-76%)
Community pharmacist	55 (67%, 57%-77%)	42 (51%, 40%-62%)	9 (11%, 4%-18%)
Hospital nurse	54 (66%, 56%-76%)	50 (61%, 50%-72%)	17 (21%, 12%-30%)
Other specialist physician	52 (63%, 53%-74%)	47 (57%, 47%-68%)	21 (26%, 16%-35%)
Community nurse	46 (56%, 45%-67%)	36 (44%, 33%-55%)	10 (12%, 5%-19%)
Surgeon	40 (49%, 38%-60%)	32 (39%, 29%-50%)	15 (18%, 10%-27%)
Psychiatrist or psychologist	31 (38%, 27%-48%)	25 (30%, 21%-41%)	19 (23%, 14%-32%)
Palliative care physician	25 (30%, 21%-40%)	19 (23%, 14%-32%)	13 (16%, 8%-24%)

[a]HCP: health care professional.

[b]HRQoL: health-related quality of life.

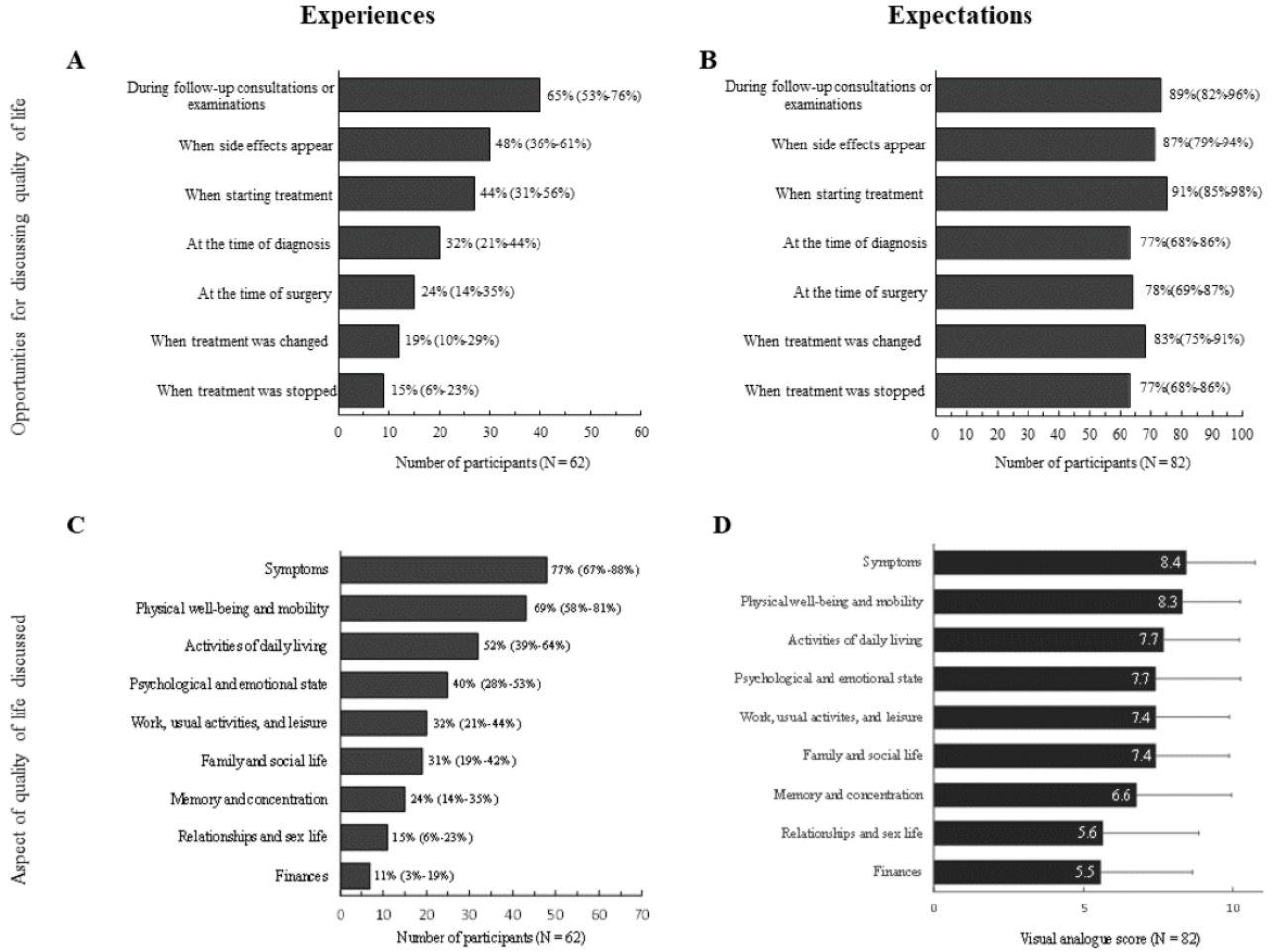
[c]Patients stated that they had discussed HRQoL at least once with indicated HCPs.

Of 62 patients, 8 (13%) reported that they had discussed their HRQoL in settings other than medical consultations, such as with patient support groups, discussion groups, or patient groups organized by a nurse.

Opportunities for Discussing QoL

QoL was most frequently discussed during follow-up consultations (40/62, 65%) and less frequently at the time the diagnosis was made (27/62, 32%). In particular, HRQoL was addressed when patients reported experiencing side effects or when a new treatment was initiated (Figure 1A). However, most of the patients considered that it was also important to discuss HRQoL at the time of diagnosis (63/82 [77%] expected vs 26/82 [32%] experienced) and to maintain a dialogue throughout their treatment, notably when starting treatment (75/82 [92%] expected vs 36/82 [44%] experienced) and when changes were made to treatment (68/82 [83%] expected vs 16 [19%] experienced) (Figure 1A,B).

Figure 1. Opportunities for discussing QoL and the aspects of QoL discussed. QoL: quality of life.



Dimensions of QoL Discussed

Multiple dimensions of HRQoL were usually discussed, with 27 of 62 (47%) patients having discussed 4 or more dimensions. The most frequently discussed dimensions were symptoms, and physical well-being and mobility (Figure 1C). These dimensions were also those that patients thought that it was important to discuss (Figure 1D). However, expectations remained high even for dimensions less frequently addressed, such as memory and concentration, relationships and sex life, or finances.

Measures for Improving Discussions of QoL

The ways that the health care team could be more attentive to HRQoL that were spontaneously cited most commonly were a better dialogue or a more personal relationship with the HCP (26 citations), more support and guidance (15 citations), and having more personalized information about the disease and treatment (14 citations). With respect to specific measures (Table 5), the most frequently selected were better follow-up of the side effects of treatment (31/82, 38%), the provision of consultations specifically devoted to HRQoL (30/82, 37%), and better coordination of care within the health team (28/82, 34%).

Table 5. Specific measures for improving dialogue about the quality of life (N=82).

Theme	Number of patients citing theme, n (%, 95% CI)
Better follow-up of side effects	31 (38%, 27%-48%)
Specific QoL[a] consultation	30 (37%, 26%-47%)
Better coordination of care	28 (34%, 24%-44%)
Therapeutic education/patient groups	23 (28%, 18%-38%)
Better training on QoL for HCPs[b]	23 (28%, 18%-38%)
Tools for discussing QoL	22 (27%, 17%-36%)
Discussion group/patient support group	14 (17%, 9%-25%)
Longer consultations	13 (16%, 8%-24%)
Systematic involvement of a psychiatrist	12 (15%, 7%-22%)
Involvement of a social worker	10 (12%, 5%-19%)
Other[c]	3 (4%, 0%-8%)

[a]QoL: quality of life.

[b]HCP: health care professional.

[c]One case each of no special needs, patient in survival stage, more resources and time for hospital staff.

Discussion

Principal Findings

The results of this study highlight the importance of discussions of HRQoL between patients with cancer treated, or previously treated, with ICIs and their HCPs throughout the treatment journey. Overall, 75 of 82 (91%) patients reported that it was quite or very important to discuss their HRQoL with an HCP. In practice, HRQoL was discussed with an HCP in the majority of cases (62/82, 76%), and most of these patients (45/82, 55%) were satisfied with the quality of the dialogue. Nevertheless, an important gap remains between patients' expectations and real-life practice, with a significant minority of patients (19/62, 31%) who were either dissatisfied with the way their HRQoL had been discussed or would have liked to have had an opportunity to discuss it. The gap is even more significant in that around half of the patients who had discussed their HRQoL with an HCP (29/62, 35%) felt that they had not been listened to sufficiently or given the chance to express themselves fully.

Many studies have emphasized the beneficial effects that internet use for health issues can have on the doctor-patient relationship, by bringing the "informed patient" to play a more active role in the care process and by facilitating communication [34,35]. Patients using the internet believe that this allows them to understand their disease and its treatment better and, to a lesser extent, helps them take better care of themselves and to participate more in decision making concerning their health [36]. Informed patients also appear to be more motivated to engage in lifestyle changes to maximize the effects of the prescribed treatment [37]. Patients participating in patient forums, such as the Carenity cancer community, are likely to be more proactive in looking for information or support and may have specific expectations for the quality of care that they receive. For this reason, they may have been more likely to initiate discussions of HRQoL than patients who do not participate in such forums. They may also have higher expectations from these discussions and thus be more frequently dissatisfied. However, these assumptions could not be evaluated in this study.

This study was conducted from the patient perspective, and it would be of interest to complement these findings with a similar survey of the importance and utility of discussing HRQoL from the perspective of the HCP. This could help identify areas of convergence between patients and HCPs, as well as understanding the gap between experiences and expectations. For example, an HCP survey could help explain why some areas of HRQoL that are considered important by patients, such as memory problems, relationships, and finances, are rarely addressed by HCPs.

The study revealed that only one-third of patients discussed HRQoL issues related to their work, daily activities, and leisure activities and only 1 in 10 discussed the impact of the cancer on their finances. Since many cancer patients treated with ICIs may achieve durable survival, these treatments may allow a more rapid return to work of cancer patients and a reduction in the amount of sick leave [28], which would be expected to be accompanied by an improvement in HRQoL. The availability of ICI therapy was quite recent at the time of the study. With a longer period of patient follow-up, it would be interesting to evaluate how HRQoL perceptions may evolve over the long term in patients treated with ICIs and in particular to compare perceptions of HRQoL between patients starting ICIs and long-term survivors previously treated with ICIs.

The study findings have identified several important but unfulfilled expectations for a more satisfying dialogue about HRQoL that are widely expressed by patients. Given the importance of monitoring QoL for the management of cancer, integrating a productive dialogue about HRQoL into routine clinical practice is essential, and this study suggests a number of ways in which such a dialogue could be improved so that patients' expectations are more fully met. Such initiatives are

all the more justified in the light of many studies that have reported significant clinical benefits associated with considering the patient's perceptions of HRQoL [7-13].

First, it would be important to broaden the discussion of HRQoL and not just focus on symptoms or side effects. Only a minority of patients discussed their emotional well-being or the impact of their cancer on their family, social life or professional life, even though they rated highly the importance of discussing these subjects. This focus on symptoms and side effects at the expense of a broader approach to HRQoL has already been emphasized in previous studies of the patient-physician dialogue in patients with advanced cancer receiving standard chemotherapy [38]. Second, the dialogue should be initiated at the time of diagnosis, rather than waiting until a patient has an issue with symptoms or treatment side effects, and continued over the course of the disease. It may be appropriate to set aside specific consultations, or at least a dedicated time during a routine consultation, to talk about HRQoL. Third, the entire care team should be involved in discussion of HRQoL. From the patient's point of view, the oncologist is the key HCP for discussing HRQoL. However, sharing information about HRQoL across the care team is important to ensure optimal coordination of care. In particular, the general practitioner is also considered an important partner for discussing QoL by patients and could thus play an active role in monitoring HRQoL over the long term.

Physician education should emphasize the need to open the discussion of HRQoL with their patients proactively and systematically. In addition, getting the patient to complete an HRQoL questionnaire before each consultation may be useful for the physician to assess any evolution of HRQoL and to identify any specific issues to be discussed. Different feasibility studies are underway to systematically collect ICI-related symptom and HRQoL data [39,40] and should provide interesting complementary information. A recent study on social media suggested that existing standard HRQoL questionnaires should be enriched with new items that are more relevant for patients treated with ICIs in their daily experience with disease and treatment [29]. Participation in discussion groups or patient support programs could also be systematically proposed. Finally, technological advances now allow monitoring of HRQoL at home through a telemedicine approach using electronic patient-reported outcomes available on applications for smartphones or computers [25,40-42].

Strengths and Limitations

Like all studies, this one had a number of strengths and limitations. The study included patients with many different types of cancer (principally lung cancer, lymphoma, and melanoma) representative of the principal cancers treated with ICIs in France. The study sample was relatively representative of the target population in terms of age, gender, geographical area, and the type of care received. No data were collected on stage, since there was a doubt as to whether this information could be reliably ascertained from the panel without medical ascertainment, since the patient may not remember and because the stage might have evolved between the time of treatment and the time of the survey. Likewise, panelists were not asked about the specific ICI prescribed, although the list of treatments was specified in the questionnaire, and treatments used by patients in the Carenity platform are not documented in the platform database. For these reasons, it was not possible to investigate the representativeness of responders further, nor to evaluate how these factors might influence perceptions of HRQoL. The number of patients was also relatively small, and patients were unlikely to be representative of all patients with cancer treated or eligible for treatment with ICIs in France. This diversity of cancer types may mask specific HRQoL issues that are important in particular forms of cancer.

Since ICIs were only approved for locally advanced or metastatic cancers at the time of the study, the patient population was at an advanced stage of disease, with one-fifth of patients having been diagnosed for at least 5 years. This implies that all patients should be at a similar stage of their disease, with a current or at least recent experience of ICI therapies. This would ensure relative homogeneity of patients. However, since it may be difficult and arbitrary for patients to distinguish their HRQoL experience with different individual treatments that were managed by the same care providers, patients were invited to describe their experiences over the whole duration of their care since diagnosis. It was thus not possible to interpret patient perceptions and expectations as relating specifically to the period of treatment with ICIs. It was nonetheless possible that recent experiences may dominate earlier ones due to a recall effect.

Conclusion

In conclusion, this study identified a gap between expectations and reality in the quality of the dialogue between patients and HCPs about HRQoL and also suggested ways to narrow this gap. Patients with cancer have a legitimate desire for a comprehensive and constructive dialogue with their physicians about their QoL, and in the case of patients receiving immunotherapy, this dialogue may be expected to continue for long periods. To meet patient expectations, the dialogue should consider all dimensions of HRQoL. A dialogue about HRQoL should be integrated into clinical practice at every step of the care pathway on a continuous basis from diagnosis to palliative care. It could be facilitated operationally by new modes of care provision, for example, offering specific consultations with an HCP dedicated to discussing HRQoL. Optimizing this dialogue should thus be a priority for physicians treating patients with cancer.

Conflicts of Interest

OW, AB, and LR are employees of Carenity, an organization that provides an online platform for patient communities, including a community of patients with cancer. FEC, AFG, and HL are employees of Bristol-Myers Squibb (BMS), purveyor of immunotherapies used in different cancers.

References

1. van Roij J, Fransen H, van de Poll-Franse L, Zijlstra M, Raijmakers N. Measuring health-related quality of life in patients with advanced cancer: a systematic review of self-administered measurement instruments. Qual Life Res 2018 Aug;27(8):1937-1955. [doi: 10.1007/s11136-018-1809-4] [Medline: 29427216]
2. Ringdal GI, Ringdal K. Testing the EORTC Quality of Life Questionnaire on cancer patients with heterogeneous diagnoses. Qual Life Res 1993 Apr;2(2):129-140. [doi: 10.1007/BF00435732] [Medline: 8518767]
3. Cella DF, Tulsky DS, Gray G, Sarafian B, Linn E, Bonomi A, et al. The Functional Assessment of Cancer Therapy scale: development and validation of the general measure. J Clin Oncol 1993 Mar;11(3):570-579. [doi: 10.1200/JCO.1993.11.3.570] [Medline: 8445433]
4. Institut National du Cancer. Plan Cancer 2014-2019: Ministère des Solidarités et de la Santé. Paris URL: https://solidarites-sante.gouv.fr/ministere/documentation-et-publications-officielles/rapports/sante/article/plan-cancer-2014-2019-4eme-rapport-au-president-de-la-republique [accessed 2021-12-06]
5. Haute Autorité de Santé. Évaluation des Médicaments: HAS-Santé. 2018. URL: https://www.has-sante.fr/upload/docs/application/pdf/2021-03/doctrine_ct.pdf [accessed 2021-12-06]
6. Kelley AS, Morrison RS. Palliative Care for the seriously ill. N Engl J Med 2015 Aug 20;373(8):747-755 [FREE Full text] [doi: 10.1056/NEJMra1404684] [Medline: 26287850]
7. Kotronoulas G, Kearney N, Maguire R, Harrow A, Di Domenico D, Croy S, et al. What is the value of the routine use of patient-reported outcome measures toward improvement of patient outcomes, processes of care, and health service outcomes in cancer care? A systematic review of controlled trials. J Clin Oncol 2014 May 10;32(14):1480-1501. [doi: 10.1200/JCO.2013.53.5948] [Medline: 24711559]
8. Greenhalgh J, Gooding K, Gibbons E, Dalkin S, Wright J, Valderas J, et al. How do patient reported outcome measures (PROMs) support clinician-patient communication and patient care? A realist synthesis. J Patient Rep Outcomes 2018 Dec;2:42 [FREE Full text] [doi: 10.1186/s41687-018-0061-6] [Medline: 30294712]
9. Velikova G, Booth L, Smith AB, Brown PM, Lynch P, Brown JM, et al. Measuring quality of life in routine oncology practice improves communication and patient well-being: a randomized controlled trial. J Clin Oncol 2004 Feb 15;22(4):714-724 [FREE Full text] [doi: 10.1200/JCO.2004.06.078] [Medline: 14966096]
10. Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. JAMA 2017 Jul 11;318(2):197-198 [FREE Full text] [doi: 10.1001/jama.2017.7156] [Medline: 28586821]
11. Basch E, Deal AM, Kris MG, Scher HI, Hudis CA, Sabbatini P, et al. Symptom monitoring with patient-reported outcomes during routine cancer treatment: a randomized controlled trial. J Clin Oncol 2016 Feb 20;34(6):557-565 [FREE Full text] [doi: 10.1200/JCO.2015.63.0830] [Medline: 26644527]
12. Denis F, Yossi S, Septans A, Charron A, Voog E, Dupuis O, et al. Improving survival in patients treated for a lung cancer using self-evaluated symptoms reported through a web application. Am J Clin Oncol 2017 Oct;40(5):464-469. [doi: 10.1097/COC.0000000000000189] [Medline: 25811297]
13. Denis F, Lethrosne C, Pourel N, Molinier O, Pointreau Y, Domont J, et al. Randomized trial comparing a web-mediated follow-up with routine surveillance in lung cancer patients. J Natl Cancer Inst 2017 Sep 01;109(9):436. [doi: 10.1093/jnci/djx029] [Medline: 28423407]
14. Hargadon KM, Johnson CE, Williams CJ. Immune checkpoint blockade therapy for cancer: an overview of FDA-approved immune checkpoint inhibitors. Int Immunopharmacol 2018 Sep;62:29-39. [doi: 10.1016/j.intimp.2018.06.001] [Medline: 29990692]
15. Postow MA, Callahan MK, Wolchok JD. Immune checkpoint blockade in cancer therapy. J Clin Oncol 2015 Jun 10;33(17):1974-1982 [FREE Full text] [doi: 10.1200/JCO.2014.59.4358] [Medline: 25605845]
16. Pons-Tostivint E, Latouche A, Vaflard P, Ricci F, Loirat D, Hescot S, et al. Comparative analysis of durable responses on immune checkpoint inhibitors versus other systemic therapies: a pooled analysis of phase III trials. JCO Precis Oncol 2019 Dec(3):1-10. [doi: 10.1200/po.18.00114]
17. Michot JM, Bigenwald C, Champiat S, Collins M, Carbonnel F, Postel-Vinay S, et al. Immune-related adverse events with immune checkpoint blockade: a comprehensive review. Eur J Cancer 2016 Feb;54:139-148. [doi: 10.1016/j.ejca.2015.11.016] [Medline: 26765102]
18. Schadendorf D, Dummer R, Hauschild A, Robert C, Hamid O, Daud A, et al. Health-related quality of life in the randomised KEYNOTE-002 study of pembrolizumab versus chemotherapy in patients with ipilimumab-refractory melanoma. Eur J Cancer 2016 Nov;67:46-54 [FREE Full text] [doi: 10.1016/j.ejca.2016.07.018] [Medline: 27596353]
19. Tykodi SS, Schadendorf D, Cella D, Reck M, Harrington K, Wagner S, et al. Patient-reported outcomes with nivolumab in advanced solid cancers. Cancer Treat Rev 2018 Nov;70:75-87 [FREE Full text] [doi: 10.1016/j.ctrv.2018.08.001] [Medline: 30125799]
20. Chouaid C, Rousseau-Bussac G, Monnet I. Qualité de vie des patients atteints de cancer du poumon traités par immunothérapie. Rev des Mal Respir Actual 2018 Oct;10(4):487-490. [doi: 10.1016/s1877-1203(18)30004-1]

21. Bordoni R, Ciardiello F, von Pawel J, Cortinovis D, Karagiannis T, Ballinger M, et al. Patient-reported outcomes in OAK: a phase III study of atezolizumab versus docetaxel in advanced non-small-cell lung cancer. Clin Lung Cancer 2018 Sep;19(5):441-449.e4 [FREE Full text] [doi: 10.1016/j.cllc.2018.05.011] [Medline: 30017645]
22. Herbst RS, Baas P, Kim D, Felip E, Pérez-Gracia JL, Han J, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. Lancet 2016 Apr 09;387(10027):1540-1550. [doi: 10.1016/S0140-6736(15)01281-7] [Medline: 26712084]
23. O'Reilly A, Hughes P, Mann J, Lai Z, Teh JJ, Mclean E, et al. An immunotherapy survivor population: health-related quality of life and toxicity in patients with metastatic melanoma treated with immune checkpoint inhibitors. Support Care Cancer 2020 Feb;28(2):561-570 [FREE Full text] [doi: 10.1007/s00520-019-04818-w] [Medline: 31089820]
24. Mamoor M, Postow MA, Lavery JA, Baxi SS, Khan N, Mao JJ, et al. Quality of life in long-term survivors of advanced melanoma treated with checkpoint inhibitors. J Immunother Cancer 2020 Mar;8(1):e000260 [FREE Full text] [doi: 10.1136/jitc-2019-000260] [Medline: 32152222]
25. Iivanainen S, Alanko T, Peltola K, Konkola T, Ekström J, Virtanen H, et al. ePROs in the follow-up of cancer patients treated with immune checkpoint inhibitors: a retrospective study. J Cancer Res Clin Oncol 2019 Mar;145(3):765-774 [FREE Full text] [doi: 10.1007/s00432-018-02835-6] [Medline: 30666409]
26. Walocko FM, Ly BC, White MS, Chen SC, Yeung H. Health-related quality of life measures and immune checkpoint inhibitors: a systematic review. J Am Acad Dermatol 2020 Apr;82(4):1004-1006 [FREE Full text] [doi: 10.1016/j.jaad.2019.10.042] [Medline: 31678470]
27. Vayr F, Montastruc M, Savall F, Despas F, Judic E, Basso M, et al. Work adjustments and employment among breast cancer survivors: a French prospective study. Support Care Cancer 2020 Jan;28(1):185-192. [doi: 10.1007/s00520-019-04799-w] [Medline: 31001691]
28. Vayr F, Savall F, Bigay-Game L, Soulat J, Chouaid C, Herin F. Lung cancer survivors and employment: a systematic review. Lung Cancer 2019 May;131:31-39. [doi: 10.1016/j.lungcan.2019.03.010] [Medline: 31027695]
29. Cotté FE, Voillot P, Bennett B, Falissard B, Tzourio C, Foulquié P, et al. Exploring the health-related quality of life of patients treated with immune checkpoint inhibitors: social media study. J Med Internet Res 2020 Sep 11;22(9):e19694 [FREE Full text] [doi: 10.2196/19694] [Medline: 32915159]
30. Faury S, Foucaud J. Health-related quality of life in cancer patients treated with immune checkpoint inhibitors: a systematic review on reporting of methods in randomized controlled trials. PLoS One 2020;15(1):e0227344 [FREE Full text] [doi: 10.1371/journal.pone.0227344] [Medline: 31978145]
31. Lamas E, Coquedano C, Bousquet C, Ferrer M, Chekroun M, Zorrilla S, et al. Patients' perception of privacy of personal data, shared in online communities: are we in the presence of a paradox? Stud Health Technol Inform 2018;251:237-240. [Medline: 29968647]
32. Lambert J, Chekroun M, Gilet H, Acquadro C, Arnould B. Assessing patients' acceptance of their medication to reveal unmet needs: results from a large multi-diseases study using a patient online community. Health Qual Life Outcomes 2018 Jul 06;16(1):134 [FREE Full text] [doi: 10.1186/s12955-018-0962-3] [Medline: 29976222]
33. Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. J Natl Cancer Inst 1993 Mar 03;85(5):365-376. [doi: 10.1093/jnci/85.5.365] [Medline: 8433390]
34. Langford AT, Roberts T, Gupta J, Orellana KT, Loeb S. Impact of the internet on patient-physician communication. Eur Urol Focus 2020 May 15;6(3):440-444. [doi: 10.1016/j.euf.2019.09.012] [Medline: 31582312]
35. Thoër C. Internet: un facteur de transformation de la relation médecin-patient ? Communiquer 2013 Dec 01(10):1-24. [doi: 10.4000/communiquer.506]
36. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
37. Iverson SA, Howard KB, Penney BK. Impact of internet use on health-related behaviors and the patient-physician relationship: a survey-based study and review. J Am Osteopath Assoc 2008 Dec;108(12):699-711. [Medline: 19075034]
38. Rodriguez KL, Bayliss N, Alexander SC, Jeffreys AS, Olsen MK, Pollak KI, et al. How oncologists and their patients with advanced cancer communicate about health-related quality of life. Psychooncology 2010 May;19(5):490-499 [FREE Full text] [doi: 10.1002/pon.1579] [Medline: 19449348]
39. Mouillet G, Fritzsch J, Paget-Bailly S, Pozet A, Es-Saad I, Meurisse A, et al. Health-related quality of life assessment for patients with advanced or metastatic renal cell carcinoma treated with a tyrosine kinase inhibitor using electronic patient-reported outcomes in daily clinical practice (QUANARIE trial): study protocol. Health Qual Life Outcomes 2019 Feb 04;17(1):25 [FREE Full text] [doi: 10.1186/s12955-019-1085-1] [Medline: 30717745]
40. Benze G, Nauck F, Alt-Epping B, Gianni G, Bauknecht T, Ettl J, et al. PROutine: a feasibility study assessing surveillance of electronic patient reported outcomes and adherence via smartphone app in advanced cancer. Ann Palliat Med 2019 Apr 26;8(2):104-111 [FREE Full text] [doi: 10.21037/apm.2017.07.05] [Medline: 29156896]
41. Zini EM, Lanzola G, Quaglini S, Bossi P, Licitra L, Resteghini C. A pilot study of a smartphone-based monitoring intervention on head and neck cancer patients undergoing concurrent chemo-radiotherapy. Int J Med Inform 2019 Sep;129:404-412. [doi: 10.1016/j.ijmedinf.2019.06.004] [Medline: 31445284]

42. LeBlanc TW, Abernethy AP. Patient-reported outcomes in cancer care: hearing the patient voice at greater volume. Nat Rev Clin Oncol 2017 Dec;14(12):763-772. [doi: 10.1038/nrclinonc.2017.153] [Medline: 28975931]

Abbreviations

EORTC-QLQ: European Organisation of Research and Treatment of Cancer Quality of Life Questionnaire
FACT: Functional Assessment of Cancer Therapy
HCP: health care professional
HRQoL: health-related quality of life
ICI: immune checkpoint inhibitor
QoL: quality of life

Edited by R Kukafka; submitted 16.11.20; peer-reviewed by M Ardebil, A Naidu; comments to author 29.03.21; revised version received 12.04.21; accepted 26.10.21; published 11.01.22.

Please cite as:
Wilczynski O, Boisbouvier A, Radoszycki L, Cotté FE, Gaudin AF, Lemasson H
Integrating Quality of Life in the Care Pathway of Cancer Patients Undergoing Immunotherapy Treatment: Descriptive, Cross-sectional Survey of an Online Patient Community's Experiences and Expectations
J Med Internet Res 2022;24(1):e25792
URL: https://www.jmir.org/2022/1/e25792
doi:10.2196/25792
PMID:35014969

Original Paper

Young Adults' Use of Different Social Media Platforms for Health Information: Insights From Web-Based Conversations

Megan S C Lim[1,2,3], PhD; Annika Molenaar[2], BNut (Hons); Linda Brennan[4], PhD; Mike Reid[4], PhD; Tracy McCaffrey[2], PhD

[1]Burnet Institute, Melbourne, Australia
[2]Monash University, Melbourne, Australia
[3]University of Melbourne, Melbourne, Australia
[4]RMIT University, Melbourne, Australia

Corresponding Author:
Megan S C Lim, PhD
Burnet Institute
85 Commercial Rd
Melbourne, 3004
Australia
Phone: 61 85062403
Email: megan.lim@burnet.edu.au

Abstract

Background: Social media–delivered health promotion has demonstrated limited uptake and effectiveness among young adults. Understanding how young adults interact with existing social media platforms for health might provide insight for future health promotion interventions.

Objective: The aim of this study is to describe how young adults interact with different social media platforms for health and health information.

Methods: We used a web-based conversation methodology to collect data from 165 young adults aged 18 to 24 years. Participants participated in an extended conversation with moderators and other participants about health and social media. They were prompted to discuss how they find health information, how they use different social media platforms, and how they evaluate the trustworthiness of information. A thematic qualitative analysis was applied to the data.

Results: Young adults spent a lot of time scrolling through Facebook newsfeeds, which often resulted in seeing health-related content either from their friends, news sources, or advertisements. Some actively sought out information about specific health areas by joining groups or following relevant pages. YouTube was considered a useful source for learning about everything and was often the go-to when searching for information or advice (after Google). Young adults found the video format easy to learn from. They stated that they could identify accurate YouTube health content by cross-checking multiple videos, by feeling that the presenter was *real* and relatable, or just through instinctively judging a video's credibility. Instagram was a source of inspiration for health and wellness from those whose lives were dedicated to healthy lifestyles and fitness. Twitter, Tumblr, and Snapchat were rarely used for health information.

Conclusions: Most young adults obtain health information from social media, both actively and through passive exposure. Participants indicated looking to social media influencers for health and lifestyle inspiration and judged the credibility of sources by appearance and instinct. Health experts should try to use the channels in the way that young adults already use them; use relatable role models on Instagram and YouTube, eye-catching headlines and support groups on Facebook, and easy to follow instruction videos via YouTube.

International Registered Report Identifier (IRRID): RR2-10.1111/1747-0080.12448

(J Med Internet Res 2022;24(1):e23656) doi:10.2196/23656

KEYWORDS

social media; Facebook; Instagram; YouTube; health information; health communication; young adults

Introduction

Health Promotion Via Social Media

In recent years, there has been a proliferation of social media–delivered health promotion campaigns targeting young adults. Such campaigns use the immense popularity of social media platforms to reach young adults with messaging to improve their health behaviors and outcomes. However, despite their promise, many interventions have demonstrated poor reach or limited effectiveness. For example, a systematic review found that only 1 of 9 interventions had a statistically significant positive impact on nutritional outcomes among young adults [1]. The proportion of young adults engaging with social media interventions in this review (eg, by liking, commenting on, or sending tweets) varied from 3% to 69%, and in another review, young adult participants engaged in only between 5% and 15% of the intended interactions [2]. The uptake and acceptability of health promotion on social media has been low among young people [1,3,4].

A systematic review showed that health promotion on social media primarily comprised information dissemination and providing social support [1]. Information dissemination can be achieved through posts from the organization that may reach the target group directly, if they have chosen to follow the organization, or an organization can pay to boost the reach of posts to others in the target group (paid advertising). Social support requires relevant individuals to interact with each other. Such social networks occur naturally (eg, in private groups on Facebook), but if health experts were to set up a group, they would need existing contact with target groups or paid social media advertising.

However, young people may not want to access health information or health social support via social media. The lack of success achieved by interventions may be based on a mismatch between how young adults use social media and the way health promoters use it. Research shows that most young people prefer to source health information from websites and not from social media [5,6]. Furthermore, many perceive it as socially undesirable to discuss personal health topics such as body weight in a public forum such as social media [1,4,7,8]. Consequently, the goals of the health promoter are not congruent with the functions of the social media platform from the users' point of view.

How Young People Use Social Media for Health

Furthermore, social media is not a single entity. There are many different social media sites or platforms, each with different features, target demographics, and use practices (affordances). Previous reviews have found that almost all health promotion interventions using social media described in the literature have been delivered through Facebook, Twitter, or custom-built social networking sites [1,3,9]. However, young adults are turning away from these platforms toward image- and video-based services such as Instagram, Snapchat, and Tik Tok [10-12]. Building a novel platform for a health intervention is costly and negates the major advantages of social media, which is that young people are already there in large numbers and are familiar with platforms' appearance and functionality [9].

Understanding how young adults interact with existing social media platforms for health might provide insight for future health promotion interventions; however, few studies have made a distinction between the use of different platforms. For example, a survey of 396 college students from the United States found that information sharing was a major motivation for Facebook use, whereas Instagram and Snapchat were used more for self-expression and self-documentation [11]. Research with adolescent girls revealed that Snapchat was used for humor, Instagram was used for self-presentation, and YouTube was used for wasting time [13]. A survey of college students from the United States showed that Facebook and Snapchat were connected with real-life friends, whereas Twitter and Instagram were used to connect with strangers [14]. Ethnic Chinese adolescents living in the United States reported viewing pictures of food on Snapchat and Instagram, whereas Facebook was used more for sharing information about health and disease [15]. In 1 study, 12 women college students journaled their use of social media and described using Facebook for social support and information gathering, whereas Pinterest and Instagram were used for recipes, exercise regimens, and inspiration [16].

In addition, health information on social media platforms from health experts competes with health information from nonexperts and industry marketing [17]. Health promotion content on social media is created by health organizations and health professionals with the intention of improving the health and well-being of recipients. Fast food, alcohol, and other commercial companies are very successful in reaching young adults on social media [18,19], as are individuals such as health and lifestyle influencers, wellness gurus, and fitspiration models [17,20,21]. In the social media era, health and science experts are often less highly regarded than celebrities and social media influencers for health and lifestyle advice [22,23]. The proliferation of health information on social media means it is difficult for laypeople to differentiate an evidence-based message from misinformation in an environment where everyone appears to have expertise [24,25]. A recent review found that the credibility of social media health information is affected by language used, perceived expertise, and bandwagon cues, such as the number of likes a post receives [26]. Most studies have examined Facebook and Twitter, whereas trustworthiness on Instagram and YouTube has been studied less frequently [26,27]. The research shows that credibility is judged differently on the 2 platforms; for example, personalized language was an effective way to increase credibility on Facebook, but depersonalized tweets were more credible on Twitter [27,28].

Although patterns emerge from these studies, there is a need for further understanding of how young adults use different platforms to seek out health information and whether they trust this information. This can be used to improve our delivery of health promotion on social media and help us understand how social media can support public health campaigns. We hypothesized that young adults would use different social media platforms in different ways and sought to determine what these differences were.

Objective

The aim of this study is to describe how young adults interact with different social media platforms for health and health information.

Methods

Design

This study used qualitative data from phase 1a of the Communicating Health project [29]. The aim of the project was to understand the use of social media by young adults to engage with health-related information; improve the effectiveness of social media strategies to motivate, engage, and retain young adults in interventions to reduce the risk of obesity; and identify and disseminate effective ways to deliver these interventions via social media.

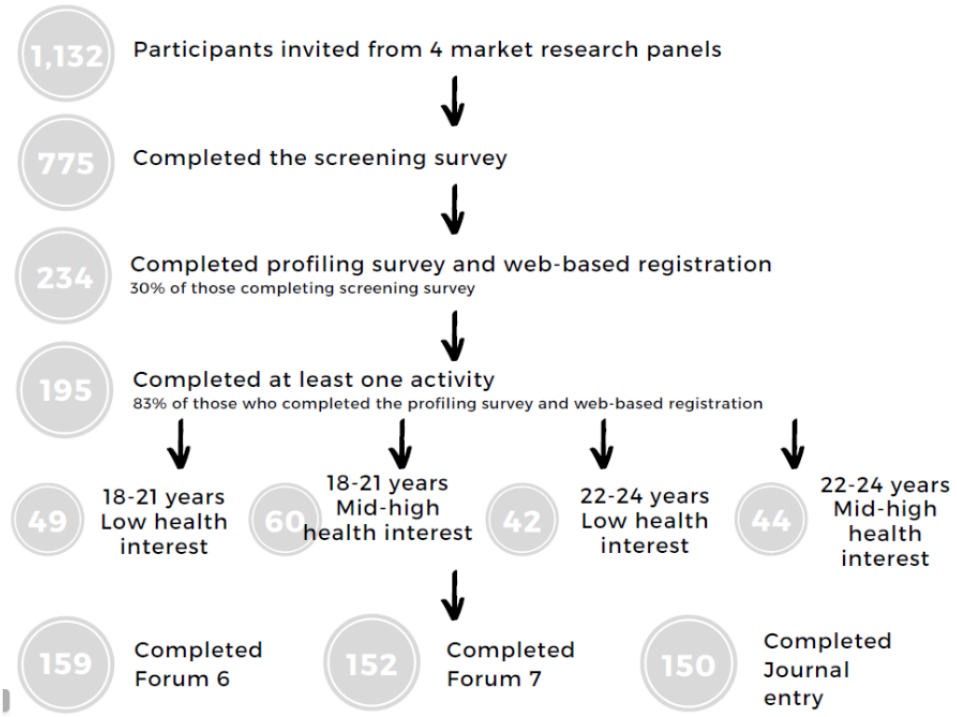
Phase 1a of Communicating Health used a web-based conversation methodology to collect data from young adults. This methodology can generate rich insights from participants using the principles of digital ethnography—a process of direct and sustained digital engagements with participants in the context of their daily lives [30]. Participants in web-based communities converse over time and respond to questions from group moderators. The conversations lasted 4 weeks and covered a series of social media–, health-, and eating-related topics. They were hosted and moderated by an independent marketing research agency and took place in a specifically created digital lounge room.

Participant Recruitment

A research society–certified field house was used to recruit participants from 3 different International Organization for Standardization accredited panels [31,32]. Young adults who had previously consented to participate in the research with 1 of 3 market and social research panels were invited to participate in this study. Participants were also able to invite their friends who were then required to undergo the same screening and profiling process. Participants who contributed to all of the forums, challenges, and polls received a gift voucher for Aus $100 (US $71), and the 20 participants who gave the most exhaustive contributions received an additional remuneration of Aus $100 (US $71).

Figure 1 summarizes participant flow. Research panels invited participants by email to complete a screening survey (completed by 775 participants). Young adults (aged 18-24 years) using social media at least twice daily and living in Australia were eligible to participate. Eligible participants were then invited to register themselves on the digital lounge room used to host web-based conversations and complete a profiling survey. The profiling survey included self-reported weight and height, demographic information, social media use, and interest in health (completed by 234 participants). This was used to assign individuals to 4 approximately equal web-based communities based on age, 18-21 years and 22-24 years, and interest in health, low or mid to high. The classification of interest of health was based on the median split from the following question completed in the profiling survey: “On a scale of 1-7 where 1 means ‘Strongly Disagree’ and 7 means ‘Strongly Agree’, please indicate how strongly you agree with the following statement—'I take an active interest in my health’.” These groups were not analyzed separately but were used so that participants would be placed with those of more similar viewpoints. Finally, 195 people participated in at least one activity in the web-based conversations.

Figure 1. Participant flow through screening and research.



Procedure

The web-based conversations (data collection) began on May 10, 2017, and the website remained active until June 6, 2017. Participants accessed the web-based community by logging in and creating a username. They were known to moderators and other participants only by this self-selected username. Participants then joined in an extended conversation about health and social media, in which they could respond to the insights of other participants and to prompts guided by the web-based moderators. Moderators asked for follow-up and clarification of participant responses and gave minimal prompts beyond the opening questions. The moderators included both a female moderator (Bachelor of Arts Psych Social and Master of Arts Applied Social Research) and a male moderator (Master of Management Marketing and Finance) with long-standing experience in market research.

There were 20 different forum discussions, in addition to 3 polls, 2 challenges, and 1 journal entry. For this manuscript, data in the form of text responses and uploaded images from the journal entry and 2 of the forums were analyzed (Textbox 1). These were chosen because of their specific discussion on the use of social media to source health information.

Textbox 1. Discussion prompts for discussion forums included in analysis.

Discussion prompts

- Forum 6—my sources of health information
  - "Now we'd like to know about the sources of information that we use in relation to finding out about health and healthy lifestyles, so post here anything that you have accessed/or access currently and in the process, tell us what's so interesting about them. If you can't think of any, that's fine, just say so. If you come across some new ones in the course of this community, please post them here too."
- Forum 7—seeking health information
  - "Imagine you're looking to find out a bit more information about health (you choose which topic you want to research): where are you going? Please step us through your search: which pages you went to, how you came across them, what you found interesting, what you found frustrating...Did anything make you trust what you saw? How and why? What would make you distrust the source? (i.e. what sends the bullshit alarm ringing?) Please share any screenshots of the websites/social media pages/profiles, links to them and any other visuals you might have around your search!"
- Journal (my use of social media and internet today)
  - "In this 'journal', we want you to record your daily use of social media and the internet at least 4 times across the three weeks of this community."
  - "Please list below the social media sites you have used, how often and for how long today (you can use yesterday if it's still early in the day today). Don't forget to tell us over the course of four different days!"
  - "Did you notice any ad, product, company or brand...? Tell us more about what made you notice them..."
  - "Beyond the everyday topics, what was discussed amongst your friends? E.g. my friends post about the politics of the day or about ads that are controversial or things like that...Did you engage in that conversation? How?"

Analysis

Analysis was conducted using NVivo 11 (QSR International). Following data familiarization, analysis commenced with a deductive phase to test the hypothesis that young adults would seek and be exposed to health information differently on different web-based platforms. To do this, forum and journal data were sorted by the platform types that were mentioned to create multiple sets of data, specifically Facebook, YouTube, Instagram, Pinterest, Snapchat, Twitter, Tumblr, and other platforms. Both forums and all journal entries were grouped together based on the different platforms in this stage for collective interpretation. Then, each data set was analyzed separately using an inductive approach to identify themes relating to how each platform was used for health information. Text was coded manually, and codes were grouped into themes. Data and subthemes were presented grouped under each platform, and each platform was given a theme title to best convey its distinct attributes. Quotes were presented verbatim, with identifying information redacted, and may include spelling errors.

Investigator triangulation was used to enhance the rigor of the analysis where coding of all forum responses and development of initial themes was conducted independently by 2 authors before coming together to discuss and come to a consensus on final themes [33]. The authors who analyzed the data included MSCL, a public health researcher (Bachelor of Biomedical Science, PhD), and AM, a nutrition scientist (Bachelor of Nutrition Science).

Descriptive quantitative findings from the profiling survey relating to sociodemographic variables, social media use, sourcing health information on the web, and trust in health information on the web were also presented.

Ethics and Funding

This study received ethics approval from the Royal Melbourne Institute of Technology Business College Human Ethics Advisory Network (project number 20489) and Monash University Human Research Ethics Committee (project number: 7807). The Communicating Health project was funded through an Australian National Health and Medical Research Council Targeted Call for Research into Engaging and Retaining Young

Adults in Interventions to Improve Eating Behaviours and Health Outcomes (GNT1115496).

Results

Participant Characteristics and Social Media Use

A total of 165 participants engaged in at least one of the forums were included in this analysis. Owing to the forums and journal entry being released in different weeks, there were different completion rates for each (forum 6: 159/165, 96.4%; forum 7: 152/165, 92.1%; and journal entry: 150/165, 90.9%). Participant characteristics for those who completed at least one of these activities is shown in Table 1.

In total, 89.1% (147/165) of the participants said they used social media 3 or more times a day; 10.9%, (18/165) twice a day. Facebook, YouTube, Snapchat, and Instagram were all used by most participants and Twitter, Pinterest, and Tumblr, by some (Figure 2). Overall, 67.8% (112/165) used social media to talk about or learn about health. Among users, approximately half used Facebook, YouTube, Instagram, and Pinterest for health reasons, but some of the users used Snapchat, Twitter, or Tumblr for health purposes.

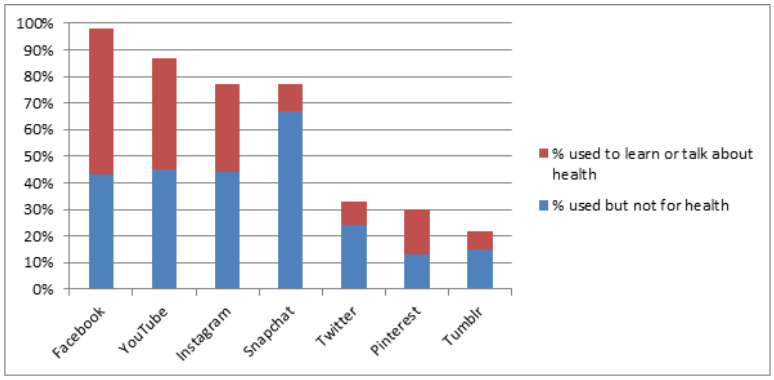
In polls, participants generally agreed that they could easily find information on the web to help them be healthy, with a median score of 6 out of 7 (IQR 5-6). They also reported a fairly high level of trust in the information they found on the internet, with a median score of 5 out of 7 (IQR 4-6).

Table 1. Participant demographics (N=165).

Characteristics	Values, n (%)
Gender	
Female	99 (60)
Male	65 (39.4)
Nonbinary or gender queer	1 (0.6)
Age (years)	
18-21	91 (55.2)
22-24	74 (44.8)
Language spoken at home	
English	121 (73.3)
Other	44 (26.7)
Currently studying	
Yes	110 (66.7)
No	55 (33.3)
BMI (kg/m^2)[a]	
Underweight (<18.5)	17 (10.3)
Healthy weight (18.5-25)	88 (53.3)
Overweight (25-30)	36 (21.8)
Obese (>30)	24 (14.5)
Live with parents	
Yes	81 (49.1)
No	84 (50.9)
Location of residence	
Metropolitan	135 (81.8)
Rural	30 (18.2)

[a]BMI by self-reported weight and height, with World Health Organization cut-offs applied.

Figure 2. Proportion using different social media channels, including those who used to talk about health (N=165).



Use of Social Media Channels

Social Media Channels

Different channels were used differently for health purposes. Even when participants described following the same companies or individuals across multiple platforms so they could see all their available content, they tended to use each platform differently:

> *Since last year I've been following [influencer – name redacted]. I discovered her via Pinterest and then subscribed to her email list in which I get sent monthly workout calendars for each day of the month that are linked to her online youtube channel that shows you the workout. I also access her website for healthy recipes and nutritional infographics relating to clean eating, antioxidants and superfoods, etc.* [Female participant, aged 18-21 years, low health interest]

Facebook (Social and Scrolling)

Many young adults described Facebook as a boredom killer. This meant they spent a lot of time scrolling through their newsfeeds, which often resulted in seeing health-related content. These were often in the form of news articles that appeared on their newsfeed or advertisements and sponsored posts, particularly from health-related retail companies. Many participants described tuning out or not paying attention to advertisements on their newsfeeds; some people believed they naturally do not notice advertisements as they are used to ignoring them. Clickbait was discussed in both positive (attention-grabbing, making them want to read the information) and negative ways (reduces the credibility of the information source and decreases the likelihood that they will click on the source, as it would likely only have sensationalized information):

> *Facebook: I only use it when I'm bored/to message friends, so about two hours of usage a day.* [Male participant, aged 22-24 years, high health interest]

> *I tend not to seek out information about having a healthy lifestyle, but I do occasionally click on articles on my facebook feed about health.* [Female participant, aged 18-21 years, high health interest]

> *I noticed a facebook ad for skill share, my local gyms, the digital garage training by google and Specsavers. I noticed them as I scrolled through my facebook feed because you sometimes forget that there are ads within facebook and assume that every post is from something that you have "liked."* [Female participant, aged 18-21 years, high health interest]

On the other hand, many actively sought out information about specific health areas they were interested in, usually through joining groups or following relevant pages. A few also mentioned finding relevant local classes and events through Facebook. Young adults were able to find groups that aligned with their specific interests and outlook, including those that have shared similar life experiences to them, such as having the same medical condition. These groups were viewed highly and often were used as the first point of call for information related to many health issues. The participants did not often discuss how they found these pages or why they specifically chose to follow those pages or groups that they did:

> *[Magazine - name redacted] have a Facebook page where they post articles about what are the right exercises to do for a specific muscle, reading it helps me understand my body much more easily.* [Male participant, aged 18-21 years, high health interest]

> *Oh god, SO many facebook groups! I've started following a lot of recipe pages and I'm in a great group called [Facebook group - name redacted] that's very anti "health-veganism". That's not to say that people don't share awesome healthy recipes that they've created, but more that it's not about health-shaming or policing what people eat, more about helping them to lead happy lives eating in a way they feel is ethically sound. it's great because so many vegan pages are really elitist, and this one makes a point that veganism can be fun and delicious and interesting without being judgemental or condescending.* [Nonbinary gender participant, aged 22-24 years, high health interest]

> *For discussion/forums, if it a personal issue I will then search for a FB group to join. You definitely learn a lot more from others personal experiences with a medical issue.* [Female participant, aged 22-24 years, high health interest]

The participants also described health-related interactions with their friends through Facebook. For example, they described being influenced by friends sharing information about their health behaviors and lifestyle. They also witnessed or engaged in discussions among friends about health topics, such as the

gym and healthy eating, either publicly on Facebook or in Facebook private messenger groups. People noted that friends shared positive stories about successful weight loss:

> *Got persuaded to get healthier through facebook where friends posted their workout routines and summer bodies haha, it got me thinking how i would feel if i workout and get better too.* [Male participant, aged 22-24 years, low health interest]

Friends also communicated with each other by sharing or tagging advertising content. Some participants would base their willingness to buy health products that companies were selling on their friends' experiences or experiences and reviews from other people. They would often trust these products if they received positive feedback:

> *My friendship circle basically communicates via sharing, tagging and liking each other in these posts...for example my bestie is obsessed with the latest [supermarket - name redacted] ad with the pasta sauces and will message me when she sees it usually just saying "PASTA AD" it's hilarious.* [Female participant, aged 18-21 years, low interest in health]

> *I noticed some of my friend's discussing brands of protein powders and whether or not they are really effective or just a money-grabber which I thought was interesting.* [Female participant, aged 18-21 years, low health interest]

Buzzfeed, particularly the Tasty channel, was mentioned by many young adults who were predominantly exposed to this channel through Facebook. It was praised for showing delicious food and easy to follow recipe videos and tutorials. The content was described as high quality and informative:

> *Food videos, sites like Tasty and Buzzfeed Food are amazing! The food always looks so good, and they provide really simple recipes for people to try! I have so many in my saves for later! I always share those two things around, I feel like they bring a happiness or a joy to my feed that other people can enjoy and get a kick out of as well!* [Female participant, aged 18-21 years, high health interest]

Some participants stated that information on Facebook was unreliable or misleading because the information could be from anyone and does not necessarily reflect the opinions of health professionals. Information clearly provided by a health professional was considered trustworthy. Other participants described the information on Facebook as unreliable but with no specific reason as to why.

YouTube (Instruction Kit for Life)

YouTube was considered a useful source for learning about everything and was often the go-to when searching for information or advice (after Google). YouTube was something they used regularly either directly as a search engine or more passively through videos recommended by YouTube based on their previous search history. Many liked the video format that was easier to follow than reading the text. Demonstrations and tutorials in visual format also made it easier to learn new skills and check in that they were doing things properly. Work out videos were particularly popular and were cited as a quick and free alternative to gyms:

> *I get most of my health content from youtube, usually because an actual demonstration of how to do an exercise or how to cook something is really helpful, and it makes me more motivated to follow through.* [Female participant, aged 22-24 years, high health interest]

> *I really just get most of my information about health and fitness from youtube. I am one of those people who can't really learn that well by just reading off a web page. I would much rather watch a video about someone doing something, I learn a lot more that way.* [Male participant, aged 22-24 years, high health interest]

A small number of participants sought advice on specific health-related issues from health professionals on YouTube:

> *For any physiotherapy related information, I watch the video on the [channel - name redacted]. They have over 1000 videos explaining exercises to combat any muscle related injury. On a side note, I was suffering from a knee clicking issue last month and following this exercise [link] helped me overcome that issue.* [Female participant, aged 18-21 years, high health interest]

Young adults also found inspiration from people they had seen sharing their stories on YouTube. The participants liked to learn what had worked for others and what could work for them:

> *I follow a few youtubers who post vlogs and they sometimes include "what I ate in a day" or "my fitness routine" which does inspire and motivate me to maintain my healthy lifestyle.* [Female participant, aged 18-21 years, high health interest]

> *Or I'll look up blogs on YouTube to see what has worked for individual people and what they think about specific lifestyles, food recipes, weight loss methods, alternative health methods etc.* [Female participant, aged 18-21 years, high health interest]

These YouTubers were seen as open and honest, they shared large volumes of information about their diet and exercise, and provided tips and advice based on their personal experience. YouTube was seen as more personal than some of the other formats often described because they were *real people* that you could see through their videos. Trustworthiness was often based on how a YouTuber looked, their personality, likeability, first impressions, and what they perceived of the person. Videos with high production value that looked professional and included a real person talking about the topic were considered more trustworthy compared with a robotic computer-generated voice and stock images:

> *I tend to scroll past all the "fitspo" looking bloggers - i don't have a lot of trust in them because often they are also incorporating surgical enhancements to achieve their physiques, and also are often pushing particular products, or just lack real knowledge. I*

> *came across [influencer - name redacted] videos, including this one: [link] I trusted her more because she didn't look "plastic" -she came across as someone from the industry who is knowledgeable about fitness, rather than someone who got into it because it is cool.* [Female participant, aged 22-24 years, high health interest]
>
> *The "bullshit alarm" rings when the voiceover is done by a robotic voice and stock Google Images are used. I would be sure to double check the facts.* [Male participant, aged 18-21 years, low health interest]

To gather reliable information, some described watching multiple videos to cross-check their facts. Some people described that they could not always find the information they needed on YouTube and might try a subsequent source such as Google. Participants would sometimes look at YouTubers with lots of subscribers or videos with lots of views as trustworthy. Fake sounding or overdramatic information and the inclusion of a large amount of apparent product sponsorships were also seen as untrustworthy. Many participants felt that they could just instinctively determine whether the information was correct:

> *Things that make me trust what I see is stories of other people that have used that information and it working for them and statistics are a good way to get my attention. I like the website WebMD and youtube as a way to visually see the effects it has on people e.g. diets. What I find frustrating is when you're searching for answers to your problems but all the company is doing is promoting themselves.* [Female participant, aged 18-21 years, low health interest]
>
> *I would begin by looking through you tube and trying to find a few sources that look credible. I would look at the length of the video, the amount of views it has and the amount of likes to dislikes ratio in order to gauge the reliability of the source...I would then cross check the information between different videos in order to gauge how true the information is. If the information is outlandish or just does not seem correct then I would continue further research.* [Male participant, aged 22-24 years, high health interest]

Many participants described tuning out or not paying attention to advertisements at the beginning of YouTube videos; some people believed they naturally do not notice advertisements as they are used to ignoring them. Using applications that block advertisements, such as *adblocker*, was also very common.

Instagram (Inspirational Appearances)

Instagram was primarily described as a source of inspiration for health and wellness. Particularly inspiring were those whose lives were dedicated to healthy lifestyles, nutrition, and fitness, including celebrities, influencers, models, and personal trainers:

> *I follow some models and yogis who generally inspire me to practice more and eat more acai bowls.* [Female participant, aged 18-21 years, low health interest]

The focus was solely on appearance. For food, clean eating and *food porn* were frequently mentioned, and concepts such as clean eating were equated with health. Most of those discussing using Instagram for health information were women, who followed other women and aspired to be like them or sought to emulate aspects of their idealized lifestyles. There was a clear emphasis placed on the ideal attractive female body, where a *good body* was equated to health:

> *If I go on Instagram, I often see fit and pretty models that make me want to start being healthy.* [Female participant, aged 18-21 years, high health interest]
>
> *I continue to visit this model's Instagram page because I trust the information she provides, since it is her job to look her best all the time and many of her posts are not sponsored but rather candid snippets of her lifestyle.* [Female participant, aged 18-21 years, high health interest]

Advertising was also prevalent on this platform, including both overt advertising and the promotion of products by Instagram influencers or celebrities. Advertising content was often hidden among Instagram posts on their feeds and was not always distinct from other content:

> *Also checked instagram a few times i have noticed a lot of celebrities or "instagram famous" people advertising all different products including weightloss shakes and teas although it isn't exactly an ad it is its clear that they have been paid to post about it.* [Female participant, aged 22-24 years, low health interest]
>
> *On instagram I scroll top to bottom, and get confused when a post shows up from someone I don't follow: until I realise its an ad. More likely to actually watch videos or something that tricks me into thinking it belongs on my timeline if that makes sense: if its clearly out of place, I'm more likely to ignore completely.* [Female participant, aged 18-21 years, low health interest]

Instagram was often mentioned among other web-based and social media sources, mostly not as an individual's only source or main source of health-related information.

Pinterest (Curating Healthy Ideas)

A smaller group used Pinterest for a range of health reasons, but they were quite enthusiastic about it. Pinterest was used for inspiration for new ideas and tips, especially recipes and workouts. People also shared their own content to get involved in conversations. Visual presentations, including photos and infographics, were considered appealing and engaging. Pinterest was mainly used as a search engine and was useful because of its ability to create *boards* to save information and ideas for use later:

> *I do also look at health content on Pinterest - being so visual it is really engaging, and when it comes to recipes, a gorgeous photo makes me so much more likely to actually cook the meal! Pinterest also has a large number of people who are interested in meal prep, so it's easy to find inspiration.* [Female participant, aged 22-24 years, high health interest]

> *I love looking at Pinterest and usually save on my page detox diets health routines and fitness workouts on my page. I find it quite interesting and fun because they seem so simple on pinterest with their diagrams and quick and easy recipes. I think their so convienent.* [Female participant, aged 22-24 years, high health interest]

Twitter, Tumblr, and Snapchat (Not About Health)

These platforms were used by young adults but rarely for sourcing or discussing health information. A few participants mentioned following health-related personalities or companies such as *Taste* or *The Food Network* on Snapchat. Otherwise, Snapchat was used to chat with friends; Twitter was used to follow news, celebrities, and sports; and Tumblr, for memes.

Discussion

Principal Findings

Young people in this study reported that they easily found health information on the web and frequently used Facebook, YouTube, and Instagram for sourcing health information. These platforms were used differently: Facebook was used to scroll through and kill time, YouTube was used to deliberately search for advice and instructions, and Instagram was used as a source of inspiration. Health information was usually found by accident on Facebook, through advertisements or friends or by joining groups or following pages related to specific health topics of personal interest. On YouTube, health content was searched for deliberately, with multiple videos being checked to compare information given; in many cases, the quality of the information was judged based on the perceived trustworthiness of the YouTuber personality presenting it. Instagram-based health information was driven almost completely by fitness and wellness personalities. Previous health promotion using social media has relied on information dissemination and social support [1]; however, the young people in our study were more interested in using social media for inspiration and operational instruction.

Trustworthiness of Health Information on Social Media

The rise of social media as an information source has resulted in a largely unregulated body of information related to health and nutrition, making it difficult to determine what information is evidence-based [24]. Participants indicated looking to social media influencers for health and lifestyle inspiration despite these individuals often promoting information that lacks an evidence base [23,25]. The often unrealistic lifestyles promoted by these influencers and social media *fitspiration* content has potentially damaging effects on mental health and well-being, body image, and food choices [34,35]. There is a need for the public to develop media literacy skills to decipher credible sources and messages, particularly in response to the changing and evolutionary nature of a science such as nutrition and to potentially build resilience from the damaging aspects of social media [36,37].

Participants reported a fairly high level of trust in health information on the web; however, they largely relied on their instinct to determine whether information was trustworthy. Trust was predominantly based on the trustworthiness of the person and the quality of the presentation rather than on the basis of the platform or the content of the information. An individual's trustworthiness was based on their appearance and perceived *realness*. Realness was based on sharing personal stories and personal experience; for example, if someone shared that something had *worked for them personally*, it was considered evidence of effectiveness. This builds on existing media research demonstrating that the trustworthiness of the person communicating information is the most important factor in determining the trustworthiness of a message [38] and that this is not correlated with the expertise of the source [39]. The perceived credibility of different sources on social media is influenced by factors such as the type of language used, specifically the use of positively framed messages and nonbiased or opinionated messages [27]. Also affecting credibility were bandwagon heuristics such as the number of followers or likes someone on social media has and expertise heuristics such as relation to a well-known credible organization [27]. The factors that young people use to judge credibility and source authority are not consistent with recommendations; these generally suggest using or citing professionals and not testimonial style personal stories (eg, Better Health Channel [40]). Health promoters should attempt to create higher levels of personal trust in our messages, for example, by having real researchers, practitioners, or young adult ambassadors present information directly to young people and building a relationship with their audience. An experimental study also showed that nutrition professionals can enhance their trustworthiness and authenticity by using *heroic* messaging featuring positive emotions [27]. Furthermore, health experts can work with existing influencers to share evidence-based health information with their followers. There is some evidence that this can increase the reach and engagement of health campaigns [41].

Platforms for Health Information

This research also shows that health promotion campaigns need to be adapted to suit different platforms. Each platform has a different style of content and method of finding content. Health promotion efforts may benefit from using platforms in the way that young people already use them; for example, using role models on Instagram and YouTube, eye-catching headlines or support groups on Facebook, and detailed but easy to follow instructions via YouTube. We should provide messages that resonate with and use the language of our target audience. Messages should be simple and easy to understand, including a call-to-action that is achievable and communicates a tangible benefit, such as easy recipes or workouts [42]. We recommend that time should be invested in analyzing *competitor* posts to gain insights into strategies that are engaging with the target audience [17,21]. Science translation needs to be memorable, easy to visualize, and should specify when and how to act on the scientific recommendation to increase understanding and uptake of behaviors that the messages promote, by the intended target audience [43]. Overall, using the concept of real, relatable people sharing anecdotes backed by evidence is likely to be a valuable strategy for social media health communication.

Another key finding is that YouTube is likely underutilized by health promoters. After Facebook, YouTube was the second

most used platform overall and for health information by young people, but it has been relatively little studied in terms of health promotion research. Although past research mentioned YouTube as being used as a time waster [13], we found that it was also used very deliberately to seek health information and most importantly, practical instructions and guidance.

It is difficult to attract young people to a profile without a pre-existing set of followers or access to a follower group. Paid advertising is a route often used by health promotion and research to reach young people. Although many participants said that they were wary of any sponsored content, their comments revealed that they often did notice and engage with them and that advertisements were cleverly placed to be confused with natural content. Many stated that they were immune to advertising, that they just did not see it, but research shows that this sort of covert advertising on social media can be very effective [44,45]. The use of advertisement blocking software by young adults was common and may affect the reach of health promotion or social marketing campaigns that target people on social media, as these campaigns will show up as sponsored posts or advertisements unless the individuals being targeted follow the organization's page.

Limitations

The key limitation of this study is that respondents generally did not provide in-depth responses to prompts as would be possible in a traditional qualitative interview. There was some back-and-forth discussion and requests to expand on comments from moderators and other participants, but extensive probing and follow-up on comments was not possible. We would recommend that future studies using this methodology include moderators who are age-group peers of participants to increase their engagement.

Conclusions

This study provides insight into the different ways young adults use social media platforms to source health information. Health experts and health promotion practitioners could improve their engagement with young adults by using relatable and inspiring personalities, particularly on YouTube and Instagram. As new social media platforms evolve and their popularity waxes and wanes, health professionals should be cognizant of the platform used by their target audience and the types of messages that promote high engagement.

Acknowledgments

This study was funded by the National Health and Medical Research Council.

Conflicts of Interest

None declared.

References

1. Klassen KM, Douglass CH, Brennan L, Truby H, Lim MS. Social media use for nutrition outcomes in young adults: a mixed-methods systematic review. Int J Behav Nutr Phys Act 2018 Jul 24;15(1):70 [FREE Full text] [doi: 10.1186/s12966-018-0696-y] [Medline: 30041699]
2. Maher CA, Lewis LK, Ferrar K, Marshall S, De Bourdeaudhuij I, Vandelanotte C. Are health behavior change interventions that use online social networks effective? A systematic review. J Med Internet Res 2014 Feb 14;16(2):e40 [FREE Full text] [doi: 10.2196/jmir.2952] [Medline: 24550083]
3. Lim MS, Wright CJ, Carrotte ER, Pedrana AE. Reach, engagement, and effectiveness: a systematic review of evaluation methodologies used in health promotion via social networking sites. Health Promot J Austr 2016 Feb;27(3):187-197. [doi: 10.1071/HE16057] [Medline: 27719734]
4. Wright CJ, Davis A, Vella AM, McGrath-Lester R, Lim MS. Young people's needs and preferences for health resources focused on pornography and sharing of sexually explicit imagery. Public Health Res Pract 2021;31(1):e29341912.
5. Oh HJ, Lauckner C, Boehmer J, Fewins-Bliss R, Li K. Facebooking for health: an examination into the solicitation and effects of health-related social support on social networking sites. Comput Human Behav 2013 Sep;29(5):2072-2080. [doi: 10.1016/j.chb.2013.04.017]
6. Lim MS, Vella A, Sacks-Davis R, Hellard ME. Young people's comfort receiving sexual health information via social media and other sources. Int J STD AIDS 2014 Dec;25(14):1003-1008. [doi: 10.1177/0956462414527264] [Medline: 24616114]
7. Byron P, Albury K, Evers C. "It would be weird to have that on Facebook": young people's use of social media and the risk of sharing sexual health information. Reprod Health Matters 2013 May;21(41):35-44. [doi: 10.1016/S0968-8080(13)41686-5] [Medline: 23684185]
8. Wartella E, Rideout V, Montague H, Beaudoin-Ryan L, Lauricella A. Teens, health and technology: a national survey. Media Commun 2016 Jun 16;4(3):13-23. [doi: 10.17645/mac.v4i3.515]
9. Maher C, Ryan J, Kernot J, Podsiadly J, Keenihan S. Social media and applications to health behavior. Curr Opinion Psychol 2016 Jun;9:50-55. [doi: 10.1016/j.copsyc.2015.10.021]
10. Over 1.6 million Australians already using TikTok. Roy Morgan. 2020. URL: http://www.roymorgan.com/findings/8289-launch-of-tiktok-in-australia-december-2019-202002240606# [accessed 2021-12-22]

11. Alhabash S, Ma M. A tale of four platforms: motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? Soc Media Soc 2017 Feb 01;3(1):205630511769154. [doi: 10.1177/2056305117691544]
12. Teens, social media and technology 2018. Pew Research Center. 2018. URL: https://www.pewresearch.org/internet/2018/05/31/teens-social-media-technology-2018/ [accessed 2021-12-22]
13. Leary MP, Clegg EN, Santella ME, Murray PJ, Downs JS, Olfert MD. Consumption of health-related content on social media among adolescent girls: mixed-methods pilot study. JMIR Form Res 2019 Mar 01;3(1):e11404 [FREE Full text] [doi: 10.2196/11404] [Medline: 30821696]
14. Phua J, Jin SV, Kim J. Uses and gratifications of social networking sites for bridging and bonding social capital: a comparison of Facebook, Twitter, Instagram, and Snapchat. Comput Human Behav 2017 Jul;72:115-122. [doi: 10.1016/j.chb.2017.02.041]
15. Zhang N, Teti M, Stanfield K, Campo S. Sharing for health: a study of Chinese adolescents' experiences and perspectives on using social network sites to share health information. J Transcult Nurs 2017 Jul;28(4):423-429. [doi: 10.1177/1043659616680268] [Medline: 27888264]
16. #Socialsupport for diet and physical activity via web 2.0: a qualitative study of college women. Southern Illinois University Carbondale. 2016. URL: https://1library.net/document/zkjkl7pq-socialsupport-diet-physical-activity-qualitative-study-college-women.html [accessed 2021-12-22]
17. Barklamb AM, Molenaar A, Brennan L, Evans S, Choong J, Herron E, et al. Learning the language of social media: a comparison of engagement metrics and social media strategies used by food and nutrition-related social media accounts. Nutrients 2020 Sep 16;12(9):2839 [FREE Full text] [doi: 10.3390/nu12092839] [Medline: 32948033]
18. McCloud RF, Kohler RE, Viswanath K. Cancer risk-promoting information: the communication environment of young adults. Am J Prev Med 2017 Sep;53(3S1):S63-S72 [FREE Full text] [doi: 10.1016/j.amepre.2017.03.025] [Medline: 28818248]
19. Lim MS, Hare JD, Carrotte ER, Dietze PM. An investigation of strategies used in alcohol brand marketing and alcohol-related health promotion on Facebook. Digit Health 2016;2:2055207616647305 [FREE Full text] [doi: 10.1177/2055207616647305] [Medline: 29942554]
20. Carrotte ER, Vella AM, Lim MS. Predictors of "Liking" three types of health and fitness-related content on social media: a cross-sectional study. J Med Internet Res 2015 Aug 21;17(8):e205 [FREE Full text] [doi: 10.2196/jmir.4803] [Medline: 26297689]
21. Klassen KM, Borleis ES, Brennan L, Reid M, McCaffrey TA, Lim MS. What people "like": analysis of social media strategies used by food industry brands, lifestyle brands, and health promotion organizations on Facebook and Instagram. J Med Internet Res 2018 Jun 14;20(6):e10227 [FREE Full text] [doi: 10.2196/10227] [Medline: 29903694]
22. Penders B. Why public dismissal of nutrition science makes sense: post-truth, public accountability and dietary credibility. Br Food J 2018;120(9):1953-1964 [FREE Full text] [doi: 10.1108/BFJ-10-2017-0558] [Medline: 30581197]
23. Hoffman SJ, Tan C. Biological, psychological and social processes that explain celebrities' influence on patients' health-related behaviors. Arch Public Health 2015;73(1):3 [FREE Full text] [doi: 10.1186/2049-3258-73-3] [Medline: 25973193]
24. Source credibility, expertise, and trust in health and risk messaging. Oxford Research Encyclopedias. 2017. URL: https://oxfordre.com/view/10.1093/acrefore/9780190228613.001.0001/acrefore-9780190228613-e-287 [accessed 2021-12-22]
25. Tonsaker T, Bartlett G, Trpkov C. Health information on the internet: gold mine or minefield? Can Fam Physician 2014 May;60(5):407-408 [FREE Full text] [Medline: 24828994]
26. Jenkins EL, Ilicic J, Barklamb AM, McCaffrey TA. Assessing the credibility and authenticity of social media content for applications in health communication: scoping review. J Med Internet Res 2020 Jul 23;22(7):e17296 [FREE Full text] [doi: 10.2196/17296] [Medline: 32706675]
27. Jenkins EL, Ilicic J, Molenaar A, Chin S, McCaffrey TA. Strategies to improve health communication: can health professionals be heroes? Nutrients 2020 Jun 22;12(6):1861 [FREE Full text] [doi: 10.3390/nu12061861] [Medline: 32580495]
28. Yilmaz G, Quintero Johnson JM. Tweeting Facts, Facebooking Lives: the influence of language use and modality on online source credibility. Commun Res Reports 2016 Apr 13;33(2):137-144. [doi: 10.1080/08824096.2016.1155047]
29. Lombard C, Brennan L, Reid M, Klassen KM, Palermo C, Walker T, et al. Communicating health-Optimising young adults' engagement with health messages using social media: study protocol. Nutr Diet 2018 Nov;75(5):509-519. [doi: 10.1111/1747-0080.12448] [Medline: 30009396]
30. Pink S, Horst H, Postill J, Hjorth L, Lewis T, Tacchi J. Digital Ethnography Principles and Practice. Thousand Oaks, California: SAGE Publications; 2015.
31. Questionnaire design. The Research Society. URL: https://researchsociety.com.au/ [accessed 2021-12-22]
32. Certification. ISO. URL: https://www.iso.org/certification.html [accessed 2021-12-22]
33. Patton M. Qualitative Research and Methods, 3rd edition. Thousand Oaks, California: SAGE Publications; 2002.
34. Raggatt M, Wright CJ, Carrotte E, Jenkinson R, Mulgrew K, Prichard I, et al. "I aspire to look and feel healthy like the posts convey": engagement with fitness inspiration on social media and perceptions of its influence on health and wellbeing. BMC Public Health 2018 Aug 10;18(1):1002 [FREE Full text] [doi: 10.1186/s12889-018-5930-7] [Medline: 30097034]
35. Rounsefell K, Gibson S, McLean S, Blair M, Molenaar A, Brennan L, et al. Social media, body image and food choices in healthy young adults: a mixed methods systematic review. Nutr Diet 2020 Feb;77(1):19-40 [FREE Full text] [doi: 10.1111/1747-0080.12581] [Medline: 31583837]

36. Goldberg JP, Sliwa SA. Communicating actionable nutrition messages: challenges and opportunities. Proc Nutr Soc 2011 Feb;70(1):26-37. [doi: 10.1017/S0029665110004714] [Medline: 21208498]
37. Tamplin NC, McLean SA, Paxton SJ. Social media literacy protects against the negative impact of exposure to appearance ideal social media images in young adult women but not men. Body Image 2018 Sep;26:29-37. [doi: 10.1016/j.bodyim.2018.05.003] [Medline: 29807299]
38. 'Who shared it?': how Americans decide what news to trust on social media. American Press Institute. URL: https://www.americanpressinstitute.org/publications/reports/survey-research/trust-social-media/ [accessed 2021-12-22]
39. Ma T, Atkin D. User generated content and credibility evaluation of online health information: a meta analytic study. Telematics Informatics 2017 Aug;34(5):472-486. [doi: 10.1016/j.tele.2016.09.009]
40. Finding reliable health information. Better Health. URL: https://www.betterhealth.vic.gov.au/health/servicesandsupport/finding-reliable-health-information [accessed 2021-12-22]
41. Kostygina G, Tran H, Binns S, Szczypka G, Emery S, Vallone D, et al. Boosting health campaign reach and engagement through use of social media influencers and memes. Soc Media Soc 2020 May 06;6(2):205630512091247. [doi: 10.1177/2056305120912475]
42. Molenaar A, Choi TS, Brennan L, Reid M, Lim MS, Truby H, et al. Language of health of young Australian adults: a qualitative exploration of perceptions of health, wellbeing and health promotion via online conversations. Nutrients 2020 Mar 25;12(4):887 [FREE Full text] [doi: 10.3390/nu12040887] [Medline: 32218118]
43. Ratner RK, Riis J. Communicating science-based recommendations with memorable and actionable guidelines. Proc Natl Acad Sci U S A 2014 Sep 16;111 Suppl 4:13634-13641 [FREE Full text] [doi: 10.1073/pnas.1320649111] [Medline: 25225363]
44. Wojdynski BW, Evans NJ. The Covert Advertising Recognition and Effects (CARE) model: processes of persuasion in native advertising and other masked formats. Int J Adv 2019 Sep 16;39(1):4-31. [doi: 10.1080/02650487.2019.1658438]
45. Aribarg A, Schwartz EM. Native advertising in online news: trade-offs among clicks, brand recognition, and website trustworthiness. J Market Res 2019 Nov 10;57(1):20-34. [doi: 10.1177/0022243719879711]

Edited by G Eysenbach; submitted 07.10.20; peer-reviewed by A Hunsaker, J Mayoh, T Ewais; comments to author 29.11.20; revised version received 03.02.21; accepted 19.11.21; published 18.01.22.

Please cite as:
Lim MSC, Molenaar A, Brennan L, Reid M, McCaffrey T
Young Adults' Use of Different Social Media Platforms for Health Information: Insights From Web-Based Conversations
J Med Internet Res 2022;24(1):e23656
URL: https://www.jmir.org/2022/1/e23656
doi:10.2196/23656
PMID:35040796

Original Paper

A Deep Residual U-Net Algorithm for Automatic Detection and Quantification of Ascites on Abdominopelvic Computed Tomography Images Acquired in the Emergency Department: Model Development and Validation

Hoon Ko[1*], MSc; Jimi Huh[2*], MD, PhD; Kyung Won Kim[3,4], MD, PhD; Heewon Chung[1], MSc; Yousun Ko[5], PhD; Jai Keun Kim[2], MD, PhD; Jei Hee Lee[2], MD, PhD; Jinseok Lee[1], PhD

[1]Department of Biomedical Engineering, Kyung Hee University, Yongin-si, Republic of Korea
[2]The Department of Radiology, Ajou University School of Medicine, Suwon, Republic of Korea
[3]Department of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[4]Research Institute of Radiology, Asan Medical Center, University of Ulsan College of Medicine, Seoul, Republic of Korea
[5]Biomedical Research Center, Asan Institute for Life Sciences, Asan Medical Center, Seoul, Republic of Korea
[*]these authors contributed equally

Corresponding Author:
Jinseok Lee, PhD
Department of Biomedical Engineering
Kyung Hee University
1732, Deogyeong-daero
Giheung-gu
Yongin-si, 17104
Republic of Korea
Phone: 82 312012570
Email: gonasago@khu.ac.kr

Abstract

Background: Detection and quantification of intra-abdominal free fluid (ie, ascites) on computed tomography (CT) images are essential processes for finding emergent or urgent conditions in patients. In an emergency department, automatic detection and quantification of ascites will be beneficial.

Objective: We aimed to develop an artificial intelligence (AI) algorithm for the automatic detection and quantification of ascites simultaneously using a single deep learning model (DLM).

Methods: We developed 2D DLMs based on deep residual U-Net, U-Net, bidirectional U-Net, and recurrent residual U-Net (R2U-Net) algorithms to segment areas of ascites on abdominopelvic CT images. Based on segmentation results, the DLMs detected ascites by classifying CT images into ascites images and nonascites images. The AI algorithms were trained using 6337 CT images from 160 subjects (80 with ascites and 80 without ascites) and tested using 1635 CT images from 40 subjects (20 with ascites and 20 without ascites). The performance of the AI algorithms was evaluated for diagnostic accuracy of ascites detection and for segmentation accuracy of ascites areas. Of these DLMs, we proposed an AI algorithm with the best performance.

Results: The segmentation accuracy was the highest for the deep residual U-Net model with a mean intersection over union (mIoU) value of 0.87, followed by U-Net, bidirectional U-Net, and R2U-Net models (mIoU values of 0.80, 0.77, and 0.67, respectively). The detection accuracy was the highest for the deep residual U-Net model (0.96), followed by U-Net, bidirectional U-Net, and R2U-Net models (0.90, 0.88, and 0.82, respectively). The deep residual U-Net model also achieved high sensitivity (0.96) and high specificity (0.96).

Conclusions: We propose a deep residual U-Net–based AI algorithm for automatic detection and quantification of ascites on abdominopelvic CT scans, which provides excellent performance.

(J Med Internet Res 2022;24(1):e34415) doi:10.2196/34415

KEYWORDS

ascites; computed tomography; deep residual U-Net; artificial intelligence

Introduction

Currently, computed tomography (CT) of the abdomen and pelvis continues to be the primary modality for patients who visit an emergency department for abdominal pain or trauma, especially in time-critical situations [1]. In emergency situations, immediate assessment of CT is required, but limited radiologic resources may hamper or delay the recognition of patients who need urgent intervention or surgery [2]. To overcome these challenges, the development of artificial intelligence (AI) techniques using a deep learning model (DLM) to detect critical findings on CT images might be a possible solution [3].

On abdominopelvic CT images, several findings indicate emergent or urgent conditions, including ascites (ie, intra-abdominal free fluid), free gas, abscess, and fat stranding [1]. Of these, presence of ascites is a common finding in various acute abdominal diseases and intra-abdominal organ injury [4]. In addition, quantification of ascites is also important, as the amount of free fluid may correlate with the severity of injury [5].

There has been only one study that developed a DLM to detect ascites, but that DLM did not quantify the amount of fluid. That study used a convolutional neural network (CNN) classification algorithm to discriminate CT images with fluid from CT images without fluid, which achieved 85% sensitivity and 95% specificity [3]. In contrast to that study, we attempted to develop an AI segmentation algorithm that can perform both detection of ascites as well as the quantification of the volume of ascites at the same time. A segmentation value of zero means no ascites, and segmentation values of the area of ascites can be used to quantify the exact volume of ascites. In addition, we tried to increase the detection accuracy of the AI algorithm.

Recently, several state-of-the-art DLM algorithms for segmentation of CT images have been proposed, including U-Net [6], bidirectional U-Net [7], recurrent residual U-Net (R2U-Net) [8], and a deep residual U-Net CNN [9]. U-Net is one of the deep learning networks with an encoder-decoder architecture, which employs skip connections to combine low-level feature maps from an encoder and high-level semantic feature maps from a decoder. Since U-Net allows for the use of location and context at the same time, and works well with very few training samples, it has been widely used in medical image segmentation [10-13]. In addition, variant models based on U-Net, such as bidirectional U-Net, R2U-Net, and a deep residual U-Net, have been applied to medical image segmentation.

Of these, we hypothesized that a deep residual U-Net might be the best algorithm for segmentation because it combines the strengths of residual learning and U-Net. The residual network has several advantages [14-16]. First, it accelerates the speed of training of the deep networks. Second, it requires fewer parameters by increasing the depth of the network instead of widening the network. Third, it reduces the effect of the vanishing gradient problem. Last, it provides high accuracy in network performance, especially in image classification and segmentation. However, no study has been reported that used a deep residual U-Net algorithm for the segmentation of ascites on CT images. Thus, we aimed to develop an optimized deep residual U-Net algorithm to detect and quantify ascites on CT images, along with a performance comparison with other state-of-the-art networks.

Methods

Patients

This study was approved by the institutional review board of Ajou University Hospital. Informed consent was waived. From January 1 to March 1, 2020, a total of 1055 patients visited the emergency department and had abdominopelvic CT scans performed. Of these, 205 patients had ascites detected on their CT images. After excluding 5 patients who underwent noncontrast CT only, we included 200 patients as the ascites group. Of the remaining 850 patients without ascites, we chose 200 age- and sex-matched controls using the MatchIt package (version 4.0.0) in R software (version 4.0.2; The R Foundation). From the patients in the ascites group and the control group, we randomly selected 100 patients with ascites and 100 patients without ascites for training and testing AI models.

The clinical characteristics of the patients in the control group and ascites group are summarized in Table 1. In the control group, out of 200 patients, unknown cause of abdominal pain (n=140, 70.0%) was the most common disease category with normal abdominopelvic CT. In contrast, in the ascites group, out of 200 patients, cancer (n=42, 21.0%), liver cirrhosis (n=52, 26.0%), blunt trauma (n=37, 18.5%), and infection (n=28, 14.0%) were the main causes for emergency department visits. The majority of ascites were identified in the pelvic cavity.

Table 1. Demographic and clinical data of participants in the control group and ascites group.

Variables	Control group (n=200)	Ascites group (n=200)
Demographics		
Sex, n (%)		
Female	92 (46.0)	101 (50.5)
Male	108 (54.0)	99 (49.5)
Age in years, mean (SD)	59.7 (13.8)	60.2 (15.3)
Amount of ascites, n (%)		
Large	0 (0)	92 (46.0)
Moderate	0 (0)	47 (23.5)
Small	0 (0)	61 (30.5)
Disease category, n (%)		
Cancer	14 (7.0)	42 (21.0)
Congestive heart failure	0 (0)	3 (1.5)
Liver cirrhosis	1 (0.5)	51 (25.5)
Acute liver failure	0 (0)	3 (1.5)
Infection	7 (3.5)	28 (14.0)
Blunt trauma	5 (2.5)	37 (18.5)
Postoperative status	32 (16.0)	5 (2.5)
Intestinal obstruction	1 (0.5)	10 (5.0)
Renal failure	0 (0)	10 (5.0)
Unknown cause of abdominal pain	140 (70.0)	11 (5.5)

CT Image Acquisition and Analysis

All patients underwent abdominopelvic CT scans using multichannel multidetector scanners (Somatom Definition Edge or Somatom Definition AS, Siemens Healthineers). Contrast-enhanced CT scans were obtained with intravenous injections of 100 to 150 mL of a nonionic contrast medium (Iopamiro 300, Bracco Imaging; Omnipaque 300, GE Healthcare) at a rate of 2.5 to 3 mL/s. The scan parameters were as follows: beam collimation, 0.75 mm; slice thickness, 5 mm; effective tube current–time charge, 200 to 260 mAs; and voltage, 100 to 120 kVp. In this study, we used only contrast-enhancement CT images. If there were multiphasic CT images, we chose portal venous phase CT images for AI training and validation.

An expert abdominal radiologist (JH, with 13 years' experience) selected CT slices that demonstrated ascites from the ascites group (2461 images from 100 patients). Then, the radiologist selected corresponding CT slices from the control group (5511 images from 100 patients). The radiologist created segmentation maps of ascites in the selected CT slices using ImageJ software (version 1.53j; National Institutes of Health), which served as ground-truth labels.

Training and Validation Data Set and Augmentation

Table 2 summarizes the training and testing data sets, which were randomly split with a ratio of 8:2 into a training set and a testing set, respectively, in a stratified fashion. The testing set was used only for an independent test of developed models and was never used for training and internal validation.

The training data set was then further separated for training the model (80% of the training set) and for internal validation (20% of the training set). To balance the two groups' images as well as reduce overfitting on training data, we employed image augmentation. We randomly drew the training images and applied them to the random combination of angle rotation between –10 and 10 degrees and vertical and horizontal flip. Finally, a total of 48,874 CT images were augmented: 24,437 images from patients with ascites and 24,437 images from healthy subjects.

Table 2. Summary of training and testing data sets.

Group	Training data, n (%)		Testing data, n (%)		Total, n (%)	
	Subjects (n=160)	Images (n=6337)	Subjects (n=40)	Images (n=1635)	Subjects (n=200)	Images (n=7972)
Ascites	80 (50.0)	1969 (31.1)	20 (50.0)	492 (30.1)	100 (50.0)	2461 (30.9)
Control	80 (50.0)	4368 (68.9)	20 (50.0)	1143 (69.9)	100 (50.0)	5511 (69.1)

Preprocessing

For all of the images in the training and testing data sets, we first set the abdomen window according to the Digital Imaging and Communications in Medicine (DICOM) standard, which is a 400 Hounsfield Unit (HU) window width and a 60 HU window level. Subsequently, we down-sampled the DICOM images as well as masked images from an image size of 512 × 512 pixels to 256 × 256 pixels, and we normalized the pixel values to a range between 0 and 1.

Deep Residual U-Net

We proposed the model for ascites region segmentation based on a single abdomen CT image using a deep residual U-Net algorithm. Figure 1 shows the architecture of our proposed model, which is comprised of three parts: an encoder, a bridge, and a decoder. In the encoder part, the normalized 256 × 256–pixel image as input is encoded into a denser representation. The decoding part, on the other hand, recovers the ascites region by pixel-wise categorization. The bridge part connects the encoding and decoding parts.

In this study, we used the residual learning approach to facilitate the training of deep neural networks and take advantage of the ascites segmentation performance gain in abdomen CT images. Each residual block consists of two paths. One path is the forward pass through batch normalization, activation, and convolutional layers, which are repeated twice. The other path is the skip connection. The outputs from the two paths are added as a single output. In the encoder part (ie, residual blocks 1-4), the output from the residual block is fed into both a subsequent residual block and one of the residual blocks in the decoder part (ie, residual blocks 6-9). Thus, in the decoder part, the residual block has two inputs: one from the encoder and the other from the previous residual block output. In the bridge part, another residual block (ie, residual block 5) connects the encoding part to the decoding part. In this study, we found that four residual blocks in each of the encoder and decoder parts provided the best performance in ascites segmentation. We describe our numerical results and comparisons in the Results section. For all residual blocks, we used the rectified linear unit activation function.

Table 3 summarizes the hyperparameters of the convolutional layers and the output size in each residual block. The normalized 256 × 256 × 3–pixel image as input was fed into residual block 1, where we used the two convolutional layers with 32 3 × 3–pixel kernels and a stride of 1 with zero padding. The activation map with a size of 256 × 256 × 32 pixels from residual block 1 was fed into both residual block 2 and residual block 9. In residual block 2, we used two convolutional layers with 64 3 × 3–pixel kernels and strides of 2 and 1 with zero padding. The activation map with a size of 128 × 128 × 64 pixels from residual block 2 was fed into both residual block 3 and residual block 8. In residual block 3, we used two convolutional layers with 128 3 × 3–pixel kernels and strides of 2 and 1 with zero padding. The activation map with a size of 64 × 64 × 128 pixels from residual block 3 was fed into both residual block 4 and residual block 7. In residual block 4, we used two convolutional layers with 256 3 × 3–pixel kernels and strides of 2 and 1 with zero padding. The activation map with a size of 32 × 32 × 256 pixels from residual block 4 was fed into residual block 5, where we used two convolutional layers with 512 3 × 3–pixel kernels and strides of 2 and 1 with zero padding.

The activation map with a size of 16 × 16 × 512 pixels from residual block 5 was fed into residual block 6, where the input was first up-sampled to 32 × 32 × 512 pixels. In residual block 6, we used two convolutional layers with 256 3 × 3–pixel kernels and a stride of 1 with zero padding. The activation map with a size of 32 × 32 × 256 pixels from residual block 6 was fed into residual block 7, and it was concatenated with the output from residual block 3. When the two inputs were concatenated, the output from residual block 6 was up-sampled to match the size. In residual block 7, we used two convolutional layers with 128 3 × 3–pixel kernels and a stride of 1 with zero padding. The activation map with a size of 64 × 64 × 128 pixels from residual block 7 was fed into residual block 8, and it was up-sampled and concatenated with the output from residual block 2. In residual block 8, we used two convolutional layers with 64 3 × 3–pixel kernels and a stride of 1 with zero padding. The activation map with a size of 128 × 128 × 64 pixels from residual block 8 was fed into residual block 9, and it was up-sampled and concatenated with the output from residual block 1. In residual block 9, we used two convolutional layers with 32 3 × 3–pixel kernels and a stride of 1 with zero padding.

The activation map with a size of 256 × 256 × 32 pixels was then fed into the convolutional layer with a single 1 × 1–pixel kernel and a stride of 1. The resultant activation map with a size of 256 × 256 × 1 pixels was finally fed into a sigmoid layer, which provided the pixel-wise probability of the presence or absence of ascites.

Figure 1. The architecture of our proposed model for ascites region segmentation based on a single abdomen computed tomography (CT) image. ReLU: rectified linear unit.

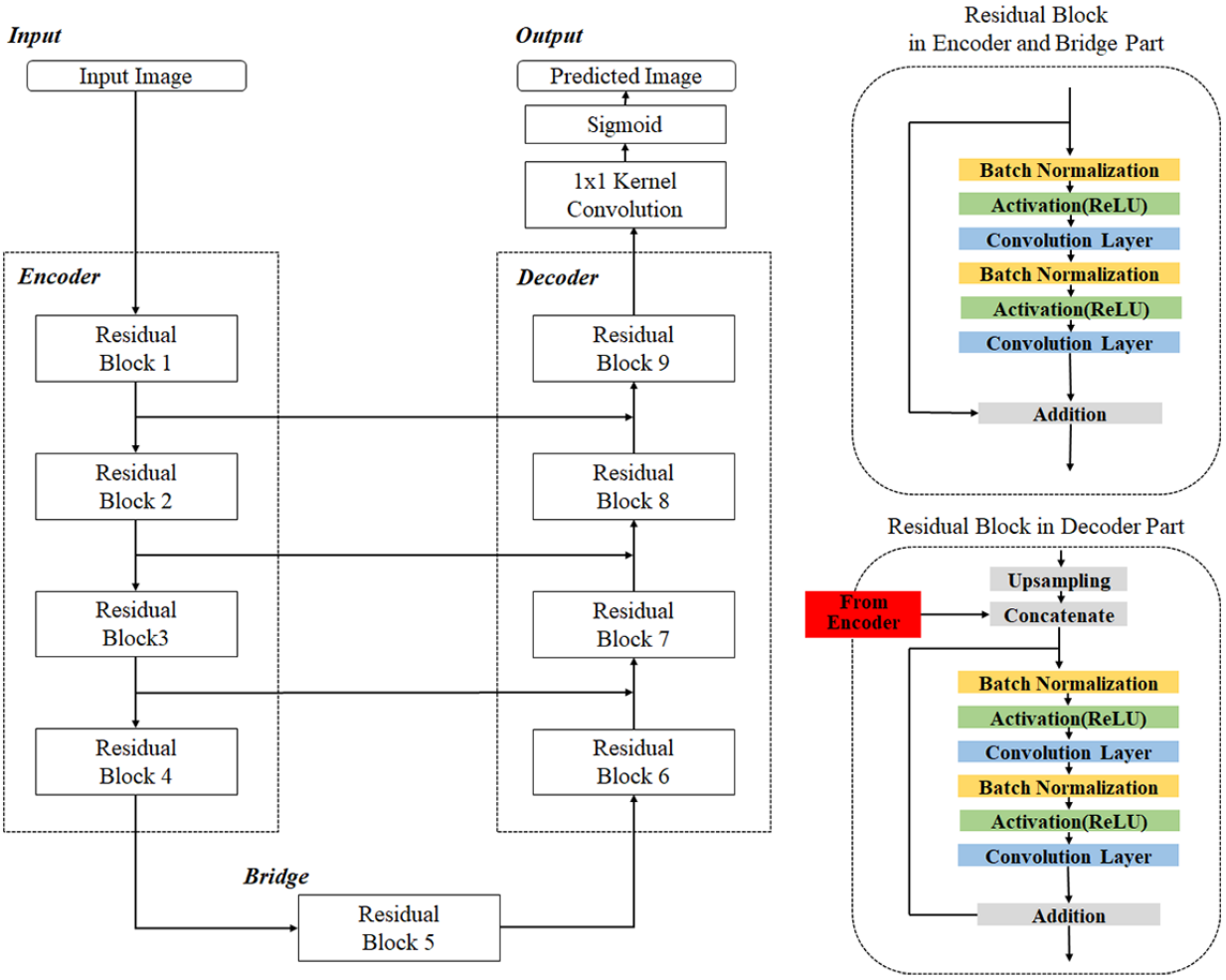


Table 3. Hyperparameters of convolutional layers according to each layer and unit level.

Model part, unit level, and layer	Kernel		Strides, n	Output size, pixels
	Filter size, pixels	Filters, n		
Input				
N/A[a]	N/A	N/A	N/A	256 × 256 × 3
Encoder				
Residual block 1				
Convolutional layer 1	3 × 3	32	1	256 × 256 × 32
Convolutional layer 2	3 × 3	32	1	256 × 256 × 32
Residual block 2				
Convolutional layer 3	3 × 3	64	2	128 × 128 × 64
Convolutional layer 4	3 × 3	64	1	128 × 128 × 64
Residual block 3				
Convolutional layer 5	3 × 3	128	2	64 × 64 × 128
Convolutional layer 6	3 × 3	128	1	64 × 64 × 128
Residual block 4				
Convolutional layer 7	3 × 3	256	2	32 × 32 × 256
Convolutional layer 8	3 × 3	256	1	32 × 32 × 256
Bridge				
Residual block 5				
Convolutional layer 9	3 × 3	512	2	16 × 16 × 512
Convolutional layer 10	3 × 3	512	1	16 × 16 × 512
Decoder				
Residual block 6				
Convolutional layer 11	3 × 3	256	1	32 × 32 × 256
Convolutional layer 12	3 × 3	256	1	32 × 32 × 256
Residual block 7				
Convolutional layer 13	3 × 3	128	1	64 × 64 × 128
Convolutional layer 14	3 × 3	128	1	64 × 64 × 128
Residual block 8				
Convolutional layer 15	3 × 3	64	1	128 × 128 × 64
Convolutional layer 16	3 × 3	64	1	128 × 128 × 64
Residual block 9				
Convolutional layer 17	3 × 3	32	1	256 × 256 × 32
Convolutional layer 18	3 × 3	32	1	256 × 256 × 32
Output				
N/A				
Convolutional layer 19	1 × 1	1	1	256 × 256 × 1
N/A				
Sigmoid layer	N/A	N/A	N/A	256 × 256 × 1

[a]N/A: not applicable; this model part did not include this parameter.

Implementation

We implemented our proposed model using the TensorFlow package (version 1.14.0), which provides a Python (version 3.6.8; Python Software Foundation) application programming interface for tensor manipulation. We also used Keras (version 2.2.4) as the official front end of TensorFlow. We trained the models with the Adam optimizer with a learning rate of 0.0001, a batch size of 16, and the loss functions of binary cross-entropy and dice loss [17] on the GeForce GTX 1080 Ti GPU (NVIDIA Corporation).

For the performance evaluation, 5-fold cross-validation was performed to confirm its generalization ability. The augmented training data set (n=48,874) was randomly shuffled and divided into five equal groups in a stratified manner. Subsequently, four groups were selected for training the model, and the remaining group was used for validation. This process was repeated five times by shifting the internal validation group. Then, we averaged the mean validation costs of the five internal validation groups according to each epoch and found the optimal epoch that provides the lowest validation cost. The testing data set was evaluated only after the model was completely trained using the training and validation data set.

Performance Evaluation

We first investigated the effect of the number of residual blocks. For the comparison, we repeated the same procedure of the 5-fold cross-validation for two to five residual blocks. For further performance comparison, we compared our proposed method with U-Net [6], bidirectional U-Net [7], and R2U-Net [8].

For the segmentation evaluation, we quantized the mean intersection over union (mIoU), which is defined as the size of the intersection divided by the size of the union. Particularly for the nonascites images, no pixel was segmented, as we quantized the value by zero. If there were no segmentation results for the nonascites image, we quantized the value by 1.

In addition to the segmentation performance, we evaluated the detection performance. If the mIoU value was equal or greater than a certain threshold value, we declared it by ascites image. For the detection performance, we plotted a receiver operating characteristic (ROC) curve and calculated the area under the ROC curve (AUROC). Subsequently, we also evaluated the sensitivity, specificity, accuracy, balanced accuracy, precision, and F1 score. More specifically, we calculated true positives (TPs), false positives (FPs), true negatives (TNs), and false negatives (FNs) and computed the following metrics:

$$\text{Sensitivity} = TP / (TP + FN) \quad \textbf{(1)}$$

$$\text{Specificity} = TN / (TN + FP) \quad \textbf{(2)}$$

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad \textbf{(3)}$$

$$\text{Balanced Accuracy} = (\text{Sensitivity} + \text{Specificity}) / 2 \quad \textbf{(4)}$$

$$\text{Precision} = TP / (TP + FP) \quad \textbf{(5)}$$

$$\text{F1 score} = 2 \times (\text{Sensitivity} \times \text{Precision}) / (\text{Sensitivity} + \text{Precision}) \quad \textbf{(6)}$$

where TP is the amount of ascites data correctly classified as ascites, TN is the amount of nonascites data correctly classified as normal, FP is the amount of nonascites data misclassified as ascites, and FN is the amount of ascites data misclassified as normal. Two abdominal radiologists (JH and KWK) also evaluated the factors influencing the performance of detection and segmentation of ascites through a systematic review of all original CT images and AI results of the testing data set.

Results

Performance in the Cross-Validation

Table 4 summarizes the cross-validation results of various AI models for ascites segmentation performance and ascites detection accuracy using mIoU and AUROC, respectively. Deep residual U-Net models with various numbers of residual blocks generally provided higher mIoU and AUROC values than any other state-of-the-art methods [6-8]. Among the deep residual U-Net models with various numbers of residual blocks, the model with four residual blocks provided the highest mIoU (0.87) for the segmentation performance and the highest AUROC (0.99) for the detection performance. The computational time for training for the deep residual U-Net model with four residual blocks and 5-fold cross-validation was 27 hours. The overall computational time for testing was 30 minutes.

Table 4. Cross-validation results for the training data set comparing the mIoU for segmentation performance and AUROC for detection across models.

Model	mIoU[a] (SD)	AUROC[b] (SD)
Deep residual U-Net (two residual blocks)	0.86 (0.03)	0.97 (0.02)
Deep residual U-Net (three residual blocks)	0.86 (0.02)	0.98 (0.01)
Deep residual U-Net (four residual blocks)	0.87 (0.02)	0.99 (0.01)
Deep residual U-Net (five residual blocks)	0.69 (0.46)	0.69 (0.01)
U-Net [6]	0.84 (0.02)	0.96 (0.01)
Bidirectional U-Net [7]	0.82 (0.01)	0.91 (0.01)
Recurrent residual U-Net [8]	0.74 (0.02)	0.90 (0.01)

[a]mIoU: mean intersection over union; this is an index of the segmentation performance.

[b]AUROC: area under the receiver operating characteristic curve; this is an index of detection accuracy.

We also investigated the effect of the number of convolutional layers in each residual block. Table 5 summarizes the cross-validation results when the number of convolutional layers changes from two to four. It shows that the deep residual U-Net model with the two convolutional layers in each residual block provided the highest values of mIoU (0.87) and AUROC (0.99), followed by three convolutional layers (mIoU=0.83 and AUROC=0.98) and four convolutional layers (mIoU=0.69 and AUROC=0.69).

Table 5. Effect of the number of convolutional layers in each residual block on cross-validation results with the training data set.

Model	mIoU[a] (SD)	AUROC[b] (SD)
Deep residual U-Net with two convolutional layers in each residual block	0.87 (0.02)	0.99 (0.01)
Deep residual U-Net with three convolutional layers in each residual block	0.83 (0.03)	0.98 (0.02)
Deep residual U-Net with four convolutional layers in each residual block	0.69 (0.02)	0.69 (0.01)

[a]mIoU: mean intersection over union; this is an index of the segmentation performance.

[b]AUROC: area under the receiver operating characteristic curve; this is an index of the detection accuracy.

Performance With the Testing Data Set

Table 6 summarizes the testing data results for segmentation performance using mIoU and detection accuracy using AUROC when the number of convolutional layers changes from two to four. Similar to the cross-validation results, these results also show that the deep residual U-Net model with four residual blocks including two convolutional layers provided the highest mIoU (0.87) and AUROC (0.96) with the isolated testing data set (n=1635).

With the two convolutional layers in each residual block, we also evaluated and compared the segmentation and detection performances. For the performance comparison, we changed the number of residual blocks from two to five and tested each model using the testing data set. Also, we tested with U-Net, bidirectional U-Net, and R2U-Net. Table 7 summarizes the performance comparison. The results also show that the deep residual U-Net with four residual blocks provided the highest mIoU and AUROC values. We also note that the deep residual U-Net with three residual blocks also provided high values of mIoU and AUROC, which were higher than any other state-of-the-art methods, indicating that the deep residual U-Net approach was more appropriate for the ascites segmentation and detection.

The representative images of ascites segmentation are presented in Figure 2. The left-hand column (A) includes the original CT images and the ground-truth masking images. Five examples of ascites segmentation results are shown using our proposed model (B) and comparing them with those using U-Net (C), bidirectional U-Net (D), and R2U-Net (E). Our proposed model correctly segmented the ascites region regardless of its pattern and size (the top four panels in column B). In addition, for the nonascites images, the segmentation results were not shown (the bottom panel in column B).

Table 8 summarizes the testing data results of detection accuracy with the metrics of sensitivity, specificity, accuracy, balanced accuracy, precision, and F1 score. The deep residual U-Net with four residual blocks provided the highest accuracy metrics: sensitivity=0.96, specificity=0.96, accuracy=0.96, balanced accuracy=0.96, precision=0.91, and F1 score=0.93. Based on these results, we proposed our deep residual U-Net with four residual blocks as an optimal AI algorithm for automatic ascites detection and segmentation on abdominopelvic CT scans.

Table 6. Effect of the number of convolutional layers in each residual block on the testing data set results for the deep residual U-Net model with four residual blocks.

Model	mIoU[a] (SD)	AUROC[b]
Deep residual U-Net with two convolutional layers in each residual block	0.87 (0.26)	0.96
Deep residual U-Net with three convolutional layers in each residual block	0.84 (0.27)	0.94
Deep residual U-Net with four convolutional layers in each residual block	0.74 (0.31)	0.72

[a]mIoU: mean intersection over union; this is an index of the segmentation performance.

[b]AUROC: area under the receiver operating characteristic curve; this is an index of the detection accuracy.

Table 7. Segmentation performance and detection accuracy of artificial intelligence models with the testing data set.

Model	mIoU[a] (SD)	AUROC[b]
Deep residual U-Net (two residual blocks)	0.81 (0.33)	0.87
Deep residual U-Net (three residual blocks)	0.86 (0.28)	0.93
Deep residual U-Net (four residual blocks)	0.87 (0.26)	0.96
Deep residual U-Net (five residual blocks)	0.70 (0.46)	0.70
U-Net [6]	0.80 (0.33)	0.90
Bidirectional U-Net [7]	0.77 (0.35)	0.86
Recurrent residual U-Net [8]	0.67 (0.41)	0.81

[a]mIoU: mean intersection over union; this is an index of the segmentation performance.

[b]AUROC: area under the receiver operating characteristic curve; this is an index of the detection accuracy.

Figure 2. Five examples of ascites segmentation results using each model. A. The original computed tomography (CT) images and the ground-truth masking images. B. Our proposed model. C. The U-Net model. D. The bidirectional U-Net model. E. The recurrent residual U-Net model. Each row represents a different example of CT images. Blue represents the ground-truth masking images, and red represents the resultant segmented images.

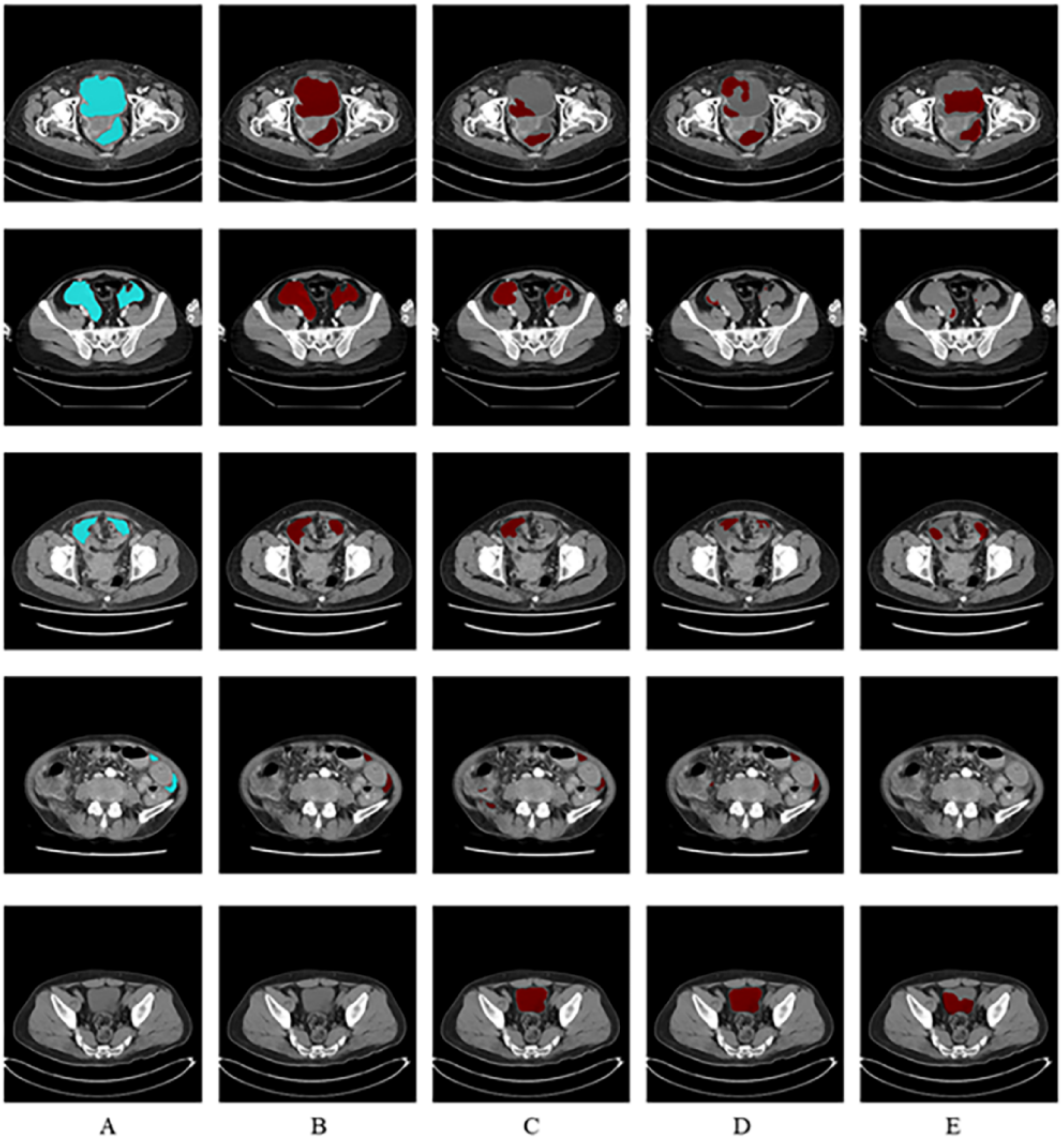


Table 8. Detection performance metrics of artificial intelligence models with the testing data set.

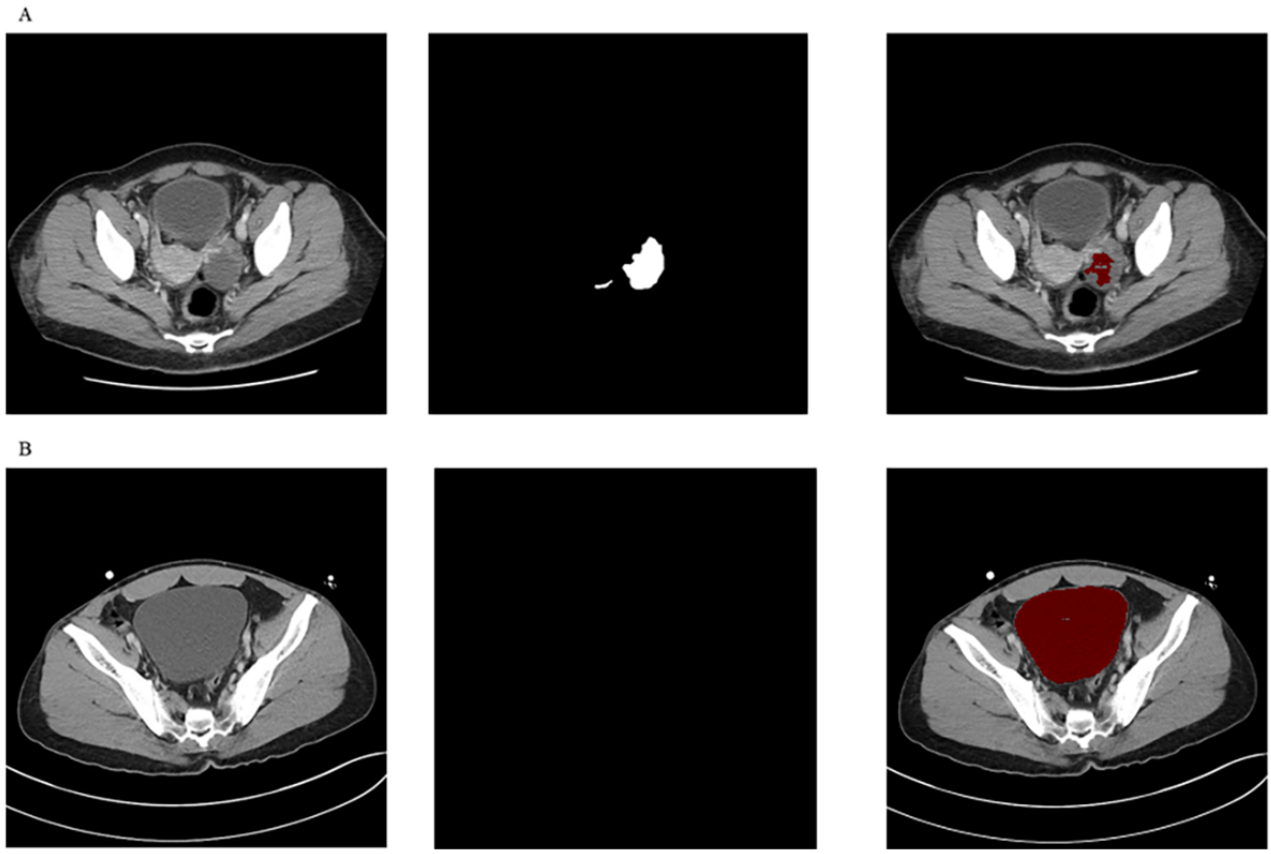
Model	Sensitivity	Specificity	Accuracy	Balanced accuracy	Precision	F1 score
U-Net [6]	0.92	0.90	0.90	0.91	0.79	0.85
Bidirectional U-Net [7]	0.94	0.86	0.88	0.90	0.74	0.83
Recurrent residual U-Net [8]	0.85	0.81	0.82	0.83	0.66	0.74
Deep residual U-Net (four residual blocks)	0.96	0.96	0.96	0.96	0.91	0.93

Factors Influencing the Performance

Through the expert review of all images in the testing data set by two radiologists (JH and KWK), there were two categories of false positive images. The AI algorithm could not differentiate between ovarian cysts of a substantial size (>3 cm in diameter) and ascites (Figure 3A). In contrast, normal physiologic ovarian cysts were correctly identified by our algorithm. The AI algorithm could not differentiate ascites from a fully distended urinary bladder (Figure 3B). However, the AI algorithm was able to differentiate ascites from a partially distended or collapsed urinary bladder.

All the false negative images showed a small amount of ascites. Two radiologists determined that all the false negative results were clinically insignificant.

Figure 3. Examples of incorrect segmentation results. The left-hand column includes the original computed tomography (CT) images, the middle column includes the ground-truth masking images, and the right-hand column includes the segmented results by our deep residual U-Net algorithm. A. In a patient with a left ovarian cyst, our artificial intelligence (AI) algorithm detected fluid within the ovarian cyst as ascites. B. In a patient with a fully distended bladder, our AI algorithm detected fluid in the bladder as ascites. Red represents the resultant segmented images.



Discussion

Principal Findings

In this study, for the first time, we developed a deep residual U-Net model for the segmentation of ascites on CT images, which provided higher accuracy compared with state-of-the-art networks, including U-Net, bidirectional U-Net, and R2U-Net. Our study results demonstrated that our AI algorithm was able to detect and quantify ascites in the abdominopelvic cavity. Our proposed algorithm was the deep residual U-Net model, which achieved 96% sensitivity, 96% specificity, and 96% accuracy for ascites detection with the testing data set. The segmentation performance was also high, with an mIoU of 0.87, when comparing the AI segmentation results and ground-truth values. However, the ground-truth values were generated by a human expert, and human error may have affected the drawing of the ascites boundaries. Thus, we believe that the AI segmentation

algorithm might be more accurate for drawing the boundary areas of ascites in general.

The deep residual U-Net model outperformed the state-of-the-art algorithms, including U-Net, bidirectional U-Net, and R2U-Net. The deep residual U-Net model combined the strengths of residual learning and U-Net architecture [9]. The network was built with residual units and has similar architecture to that of U-Net. The benefits of this model are three-fold: (1) residual units facilitate the training of deep networks, (2) the vanishing gradient problem is reduced, and (3) the rich skip connections within the network could facilitate information propagation, resulting in higher mIoU values. Integration of the residual network with standard U-Net architecture enabled us to extract robust discriminative features from input CT images.

In general, the concept of U-Net is to stitch low-level features into corresponding high-level features, thereby adding low-level texture features to high-level semantic features. Thus, U-Net with a deep layer can provide better segmentation results. However, an excessive increase in the number of network layers tends to decrease segmentation accuracy. This issue can be solved by adding a residual unit to U-Net, which can make use of the merits of the residual network [6]. A deep residual U-Net model has been used for lung segmentation in CT scans [9], joint segmentation in CT scans [18], and vulnerable plaque segmentation in optical coherence tomography images [19]. These prior studies consistently reported the high segmentation performance of a deep residual U-Net model. In addition, our proposed deep residual U-Net model has an advantage over other U-Net models, in that it requires fewer parameters compared to other tree models [6-8]. Table 9 summarizes the comparison of the number of parameters for each model. Our proposed model includes 18,855,137 weights and biases, which represents only 54.5% of the parameters from U-Net. Also, this represents only 34.0% and 78.1% of the parameters from bidirectional U-Net and R2U-Net, respectively.

Table 9. Comparison of the number of parameters for each U-Net model.

Model	Trainable parameters, n	Nontrainable parameters, n	Total parameters, n
Our proposed model	18,840,545	14,592	18,855,137
U-Net [6]	34,600,353	14,016	34,614,369
Bidirectional U-Net [7]	55,398,798	1408	55,400,197
Recurrent residual U-net [8]	24,133,013	0	24,133,013

So far, there has been only one study that developed an AI algorithm to detect ascites [3]. In that study, the authors used a CNN algorithm mainly for the classification of three abnormal CT findings of free fluid (ie, ascites), free gas, and mesenteric fat stranding. The accuracy of the CNN algorithm achieved 85% sensitivity and 95% specificity to detect ascites. In contrast, our deep residual U-Net algorithm achieved 96% sensitivity and 96% specificity for ascites detection. In addition, our deep residual U-Net algorithm also quantified the amount of ascites with high segmentation accuracy (mIoU=0.87). Thus, we believe that it is quite possible to use our proposed algorithm for ascites detection and quantification on abdominopelvic CT images in patients who visit the emergency department.

In the majority of urgent and emergent situations, clinicians should read the CT scan without radiologic support immediately after the CT scan was obtained. Getting a radiology report usually takes time, and radiologic support may not be maintained 24 hours per day in many institutions [20]. AI algorithms can help maintain radiology support in real time with high diagnostic accuracy. Our training and test data sets were unique in that CT data were obtained from patients who visited the emergency department of a tertiary care hospital, which is designated as a regional emergency medical center and a regional trauma center in Korea. Currently, we incorporated our deep residual U-Net algorithm in our radiology unit and will start further training of our algorithm in a sustainable manner.

There were false positive cases in which our AI algorithm identified fluid within organs, such as the bladder and ovarian cysts, as ascites (Figure 3). These false positive cases will decrease as we continue to train the AI algorithm. All the false negative cases showed a small amount of ascites, especially between internal organs, such as the bowels, bladder, and uterus. Further training will increase the sensitivity of the AI algorithm to detect ascites.

We adopted a 2D AI algorithm for sequential 2D image analyses rather than a 3D framework, because 3D deep learning requires higher computational power than 2D deep learning [20]. In an emergent clinical setting, a rapid AI algorithm may be preferable to a complex and slow algorithm. Our study showed that sequential 2D image analyses could provide excellent diagnostic accuracy for detecting and quantifying ascites.

Limitations and Future Work

Our study has several limitations. Firstly, we trained our model using a relatively small amount of CT data. Thus, we will establish a sustainable AI training system and train our AI algorithm using real-world CT data prospectively obtained from our emergency department. Secondly, our AI model was validated internally using a split testing data set. The testing data set was obtained from the same source as the training data set. This may raise issues of generalizability and overfitting of our model [21]. Thus, in the near future, we will validate our model using data from various institutions.

Conclusions

We propose our deep residual U-Net algorithm for the automatic detection and quantification of ascites in abdominopelvic CT scans. Our model outperformed other state-of-the-art segmentation algorithms based on U-Net, bidirectional U-Net, and R2U-Net.

Acknowledgments

This work was supported by the National Research Foundation of Korea (grant 2019R1I1A1A01060744), a grant from the Korea Health Industry Development Institute (grant HI18C1216), and the Korea Medical Device Development Fund grant, which is funded by the Government of the Republic of Korea (the Ministry of Science and ICT; the Ministry of Trade, Industry and Energy; the Ministry of Health and Welfare; and the Ministry of Food and Drug Safety) (grant KMDF_PR_20200901_0095).

Authors' Contributions

HK and HC carried out the machine learning and deep learning simulation for the hyperparameter search and modeling. JH, JKK, and JL provided the data and performed the data validation to be applied to ascites segmentation. KWK and YK validated and confirmed the simulations and helped to draft the manuscript. HK, JH, and JL wrote the initial manuscript. JL, JH, and KWK conceived the study and participated in its design and coordination. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

References

1. Levine CD, Patel UJ, Wachsberg RH, Simmons MZ, Baker SR, Cho KC. CT in patients with blunt abdominal trauma: Clinical significance of intraperitoneal fluid detected on a scan with otherwise normal findings. AJR Am J Roentgenol 1995 Jun;164(6):1381-1385. [doi: 10.2214/ajr.164.6.7754877] [Medline: 7754877]
2. Wang DC, Parry CR, Feldman M, Tomlinson G, Sarrazin J, Glanc P. Acute abdomen in the emergency department: Is CT a time-limiting factor? AJR Am J Roentgenol 2015 Dec;205(6):1222-1229. [doi: 10.2214/AJR.14.14057] [Medline: 26587929]
3. Winkel DJ, Heye T, Weikert TJ, Boll DT, Stieltjes B. Evaluation of an AI-based detection software for acute findings in abdominal computed tomography scans. Invest Radiol 2019;54(1):55-59. [doi: 10.1097/rli.0000000000000509]
4. Urban BA, Fishman EK. Tailored helical CT evaluation of acute abdomen. Radiographics 2000;20(3):725-749. [doi: 10.1148/radiographics.20.3.g00ma12725] [Medline: 10835125]
5. Sirlin CB, Casola G, Brown MA, Patel N, Bendavid EJ, Hoyt DB. Quantification of fluid on screening ultrasonography for blunt abdominal trauma: A simple scoring system to predict severity of injury. J Ultrasound Med 2001 Apr;20(4):359-364. [doi: 10.7863/jum.2001.20.4.359] [Medline: 11316314]
6. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation. In: Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. 2015 Presented at: 18th International Conference on Medical Image Computing and Computer-Assisted Intervention; October 5-9, 2015; Munich, Germany p. 234-241. [doi: 10.1007/978-3-319-24574-4_28]
7. Azad TD, Asadi-Aghbolaghi M, Fathy M, Escalera S. Bi-directional ConvLSTM U-Net with densely connected convolutions. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshop. 2019 Presented at: IEEE/CVF International Conference on Computer Vision Workshop; October 27-28, 2019; Seoul, Korea p. 406-415. [doi: 10.1109/ICCVW.2019.00052]
8. Alom MZ, Yakopcic C, Hasan M, Taha TM, Asari VK. Recurrent residual U-Net for medical image segmentation. J Med Imaging 2019 Jan 1;6(01):1. [doi: 10.1117/1.jmi.6.1.014006]
9. Khanna A, Londhe N, Gupta S, Semwal A. A deep residual U-Net convolutional neural network for automated lung segmentation in computed tomography images. Biocybern Biomed Eng 2020 Jul;40(3):1314-1327. [doi: 10.1016/j.bbe.2020.07.007]
10. Saeedizadeh N, Minaee S, Kafieh R, Yazdani S, Sonka M. COVID TV-Unet: Segmenting COVID-19 chest CT images using connectivity imposed Unet. Comput Methods Programs Biomed Update 2021;1:100007 [FREE Full text] [doi: 10.1016/j.cmpbup.2021.100007] [Medline: 34337587]
11. Schreier J, Attanasi F, Laaksonen H. A full-image deep segmenter for CT images in breast cancer radiotherapy treatment. Front Oncol 2019;9:677 [FREE Full text] [doi: 10.3389/fonc.2019.00677] [Medline: 31403032]
12. Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, et al. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Med Phys 2019 May;46(5):2157-2168 [FREE Full text] [doi: 10.1002/mp.13458] [Medline: 30810231]
13. Ma X, Hadjiiski LM, Wei J, Chan H, Cha KH, Cohan RH, et al. U-Net based deep learning bladder segmentation in CT urography. Med Phys 2019 Apr;46(4):1752-1765 [FREE Full text] [doi: 10.1002/mp.13438] [Medline: 30734932]
14. Amer A, Ye X, Janan F. ResDUnet: A deep learning-based left ventricle segmentation method for echocardiography. IEEE Access 2021;9:159755-159763. [doi: 10.1109/access.2021.3122256]
15. Singadkar G, Mahajan A, Thakur M, Talbar S. Deep deconvolutional residual network based automatic lung nodule segmentation. J Digit Imaging 2020 Jun;33(3):678-684 [FREE Full text] [doi: 10.1007/s10278-019-00301-4] [Medline: 32026218]

16. Mique Jr E, Malicdem A. Deep residual U-Net based lung image segmentation for lung disease detection. In: Proceedings of the International Conference on Information Technology and Digital Applications. 2020 Presented at: International Conference on Information Technology and Digital Applications; November 15, 2019; Yogyakarta, Indonesia URL: https://iopscience.iop.org/article/10.1088/1757-899X/803/1/012004/pdf [doi: 10.1088/1757-899X/803/1/012004]
17. Jadon S. A survey of loss functions for semantic segmentation. In: Proceedings of the IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology. 2020 Presented at: IEEE Symposium on Computational Intelligence and Bioinformatics and Computational Biology; October 27-29, 2020; Via del Mar, Chile p. 1-7. [doi: 10.1109/cibcb48159.2020.9277638]
18. Rassadin A. Deep residual 3D U-Net for joint segmentation and texture classification of nodules in lung. In: Proceedings of the 17th International Conference on Image Analysis and Recognition. 2020 Presented at: 17th International Conference on Image Analysis and Recognition; June 24-26, 2020; Póvoa de Varzim, Portugal p. 419-427. [doi: 10.1007/978-3-030-50516-5_37]
19. Li L, Jia T. Optical coherence tomography vulnerable plaque segmentation based on deep residual U-Net. Rev Cardiovasc Med 2019 Sep 30;20(3):171-177 [FREE Full text] [doi: 10.31083/j.rcm.2019.03.5201] [Medline: 31601091]
20. Ko H, Chung H, Kang WS, Kim KW, Shin Y, Kang SJ, et al. COVID-19 pneumonia diagnosis using a simple 2D deep learning framework with a single chest CT image: Model development and validation. J Med Internet Res 2020 Jun 29;22(6):e19569 [FREE Full text] [doi: 10.2196/19569] [Medline: 32568730]
21. Kim DW, Jang HY, Kim KW, Shin Y, Park SH. Design characteristics of studies reporting the performance of artificial intelligence algorithms for diagnostic analysis of medical images: Results from recently published papers. Korean J Radiol 2019 Mar;20(3):405-410 [FREE Full text] [doi: 10.3348/kjr.2019.0025] [Medline: 30799571]

Abbreviations

AI: artificial intelligence
AUROC: area under the receiver operating characteristic curve
CNN: convolutional neural network
CT: computed tomography
DICOM: Digital Imaging and Communications in Medicine
DLM: deep learning model
FN: false negative
FP: false positive
HU: Hounsfield Unit
mIoU: mean intersection over union
R2U-Net: recurrent residual U-Net
ROC: receiver operating characteristic
TN: true negative
TP: true positive

Edited by G Eysenbach; submitted 22.10.21; peer-reviewed by C Jeong, T Zhang; comments to author 15.11.21; revised version received 30.11.21; accepted 30.11.21; published 03.01.22.

Please cite as:
Ko H, Huh J, Kim KW, Chung H, Ko Y, Kim JK, Lee JH, Lee J
A Deep Residual U-Net Algorithm for Automatic Detection and Quantification of Ascites on Abdominopelvic Computed Tomography Images Acquired in the Emergency Department: Model Development and Validation
J Med Internet Res 2022;24(1):e34415
URL: https://www.jmir.org/2022/1/e34415
doi:10.2196/34415
PMID:34982041

Original Paper

Developing a Machine Learning Model to Predict Severe Chronic Obstructive Pulmonary Disease Exacerbations: Retrospective Cohort Study

Siyang Zeng[1], MS; Mehrdad Arjomandi[2,3], MD; Yao Tong[1], MSc; Zachary C Liao[1], MPH, MD; Gang Luo[1], DPhil

[1]Department of Biomedical Informatics and Medical Education, University of Washington, Seattle, WA, United States

[2]Medical Service, San Francisco Veterans Affairs Medical Center, San Francisco, CA, United States

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

Corresponding Author:
Gang Luo, DPhil
Department of Biomedical Informatics and Medical Education
University of Washington
UW Medicine South Lake Union, 850 Republican Street, Building C, Box 358047
Seattle, WA, 98195
United States
Phone: 1 206 221 4596
Fax: 1 206 221 2671
Email: gangluo@cs.wisc.edu

Abstract

Background: Chronic obstructive pulmonary disease (COPD) poses a large burden on health care. Severe COPD exacerbations require emergency department visits or inpatient stays, often cause an irreversible decline in lung function and health status, and account for 90.3% of the total medical cost related to COPD. Many severe COPD exacerbations are deemed preventable with appropriate outpatient care. Current models for predicting severe COPD exacerbations lack accuracy, making it difficult to effectively target patients at high risk for preventive care management to reduce severe COPD exacerbations and improve outcomes.

Objective: The aim of this study is to develop a more accurate model to predict severe COPD exacerbations.

Methods: We examined all patients with COPD who visited the University of Washington Medicine facilities between 2011 and 2019 and identified 278 candidate features. By performing secondary analysis on 43,576 University of Washington Medicine data instances from 2011 to 2019, we created a machine learning model to predict severe COPD exacerbations in the next year for patients with COPD.

Results: The final model had an area under the receiver operating characteristic curve of 0.866. When using the top 9.99% (752/7529) of the patients with the largest predicted risk to set the cutoff threshold for binary classification, the model gained an accuracy of 90.33% (6801/7529), a sensitivity of 56.6% (103/182), and a specificity of 91.17% (6698/7347).

Conclusions: Our model provided a more accurate prediction of severe COPD exacerbations in the next year compared with prior published models. After further improvement of its performance measures (eg, by adding features extracted from clinical notes), our model could be used in a decision support tool to guide the identification of patients with COPD and at high risk for care management to improve outcomes.

International Registered Report Identifier (IRRID): RR2-10.2196/13783

(J Med Internet Res 2022;24(1):e28953) doi:10.2196/28953

KEYWORDS

chronic obstructive pulmonary disease; machine learning; forecasting; symptom exacerbation; patient care management

Introduction

Background

In the United States, chronic obstructive pulmonary disease (COPD) affects 6.5% of adults [1] and is the fourth leading cause of death, excluding COVID-19 [2]. Each year, COPD causes 1.5 million emergency department (ED) visits, 0.7 million inpatient stays, and US $32.1 billion in total medical cost [1]. Severe COPD exacerbations are those that require ED visits or inpatient stays [3], account for 90.3% of the total medical cost related to COPD [4], and often cause irreversible decline in lung function and health status [5-10]. Many severe COPD exacerbations (eg, 47% of the inpatient stays for COPD) are deemed preventable with appropriate outpatient care [3,11] because COPD is an ambulatory care–sensitive condition [12]. A commonly used method to reduce severe COPD exacerbations is to place patients at high risk in a care management program for preventive care [13-15]. Patients at high risk can be identified prospectively using a predictive model [16]. Once a patient enters the care management program, a care manager will periodically contact the patient for health status assessment and to help coordinate health and related services. This method is adopted by many health plans, such as those in 9 of 12 metropolitan communities [13], and many health care systems. Successful care management can reduce up to 27% of the ED visits [14] and 40% of the inpatient stays [15] in patients with COPD.

However, because of limitations of resources and service capacity, only ≤3% of patients could enter a care management program [17]. Its effectiveness is upper bounded by these patients' risk levels, which are determined by how accurate the used predictive model is. Neither the stage of COPD nor having prior severe COPD exacerbations alone can predict a patient's risk level for future severe COPD exacerbations well [18,19]. Previously, researchers had built several models to predict severe COPD exacerbations in patients with COPD [20-53]. These models are inaccurate and suboptimal for use in care management because they missed more than 50% of the patients who will experience severe COPD exacerbations in the future, incorrectly projected many other patients to experience severe COPD exacerbations [20-22,53], used data unavailable in routine clinical practice [23-31,33,34,36,42-50,52], or were designed for patients who have different characteristics from typical patients with COPD [25-34]. In addition, most of these models predicted only inpatient stays for COPD. To better guide the use of care management, we need to predict both ED visits and inpatient stays for COPD, which only 2 of these models [34,36] do. In practice, once a model is deployed for care management, the prediction errors produced by the model would lead to degraded patient outcomes and unnecessary health care costs. Because of the large number of patients with COPD, even a small improvement in model accuracy coupled with appropriate preventive interventions could help improve outcomes and avoid many ED visits and inpatient stays for COPD every year.

Objective

This study aims to develop a more accurate model to predict severe COPD exacerbations in the next year in patients with COPD. To be suitable for use in care management, the model should use data available in routine clinical practice and target all patients with COPD.

Methods

Ethics Approval and Study Design

The institutional review board of the University of Washington Medicine (UWM) approved this secondary analysis study on administrative and clinical data.

Patient Population

In Washington state, the UWM is the largest academic health care system. The UWM enterprise data warehouse includes administrative and clinical data from 3 hospitals and 12 clinics. The patient cohort consisted of the patients with COPD who visited any of these facilities between 2011 and 2019. Using our prior method for identifying patients with COPD [54] that was adapted from the literature [55-58], we regarded a patient to have COPD if the patient was aged ≥40 years and met ≥1 of the 4 criteria listed in Textbox 1. When computing the data instances in any year, we excluded the patients who had no encounter at the UWM or died during that year. No other exclusion criterion was used.

Textbox 1. The 4 criteria used for identifying patients with chronic obstructive pulmonary disease.

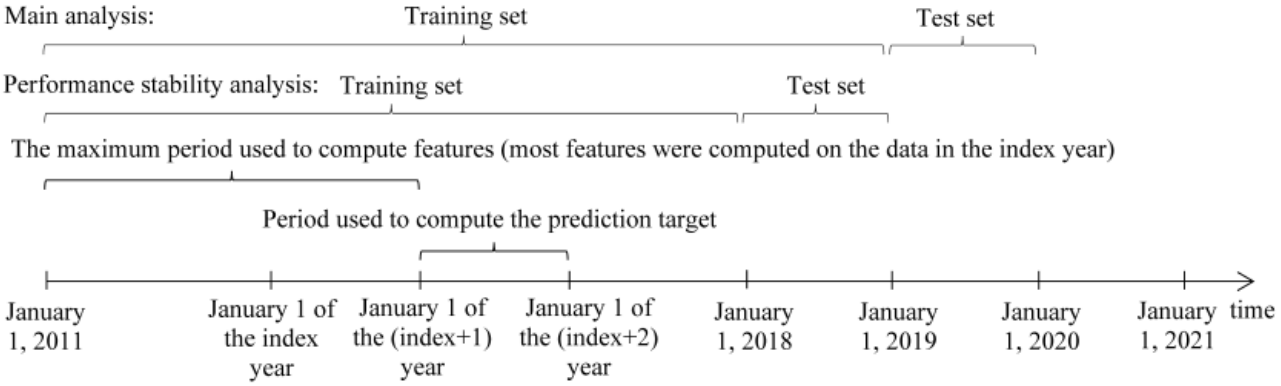
Description of each of the 4 criteria

- An outpatient visit diagnosis code of chronic obstructive pulmonary disease (International Classification of Diseases, Ninth Revision: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; International Classification of Diseases, Tenth Revision: J42, J41.8, J44.*, J43.*) followed by ≥1 prescription of long-acting muscarinic antagonist (aclidinium, glycopyrrolate, tiotropium, and umeclidinium) within 6 months
- ≥1 emergency department or ≥2 outpatient visit diagnosis codes of chronic obstructive pulmonary disease (International Classification of Diseases, Ninth Revision: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; International Classification of Diseases, Tenth Revision: J42, J41.8, J44.*, J43.*)
- ≥1 inpatient stay discharge having a principal diagnosis code of chronic obstructive pulmonary disease (International Classification of Diseases, Ninth Revision: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; International Classification of Diseases, Tenth Revision: J42, J41.8, J44.*, J43.*)
- ≥1 inpatient stay discharge having a principal diagnosis code of respiratory failure (International Classification of Diseases, Ninth Revision: 518.82, 518.81, 799.1, 518.84; International Classification of Diseases, Tenth Revision: J96.0*, J80, J96.9*, J96.2*, R09.2) and a secondary diagnosis code of acute chronic obstructive pulmonary disease exacerbation (International Classification of Diseases, Ninth Revision: 491.22, 491.21, 493.22, 493.21; International Classification of Diseases, Tenth Revision: J44.1, J44.0)

Prediction Target (Also Known as the Outcome or the Dependent Variable)

Given a patient with COPD who had ≥1 encounter at the UWM in a specific year (the index year), we used the patient's data up to the last day of the year to predict the outcome of whether the patient would experience any severe COPD exacerbation, that is, any ED visit or inpatient stay with a principal diagnosis of COPD (International Classification of Diseases, Ninth Revision: 491.22, 491.21, 491.9, 491.8, 493.2x, 492.8, 496; International Classification of Diseases, Tenth Revision: J42, J41.8, J44.*, J43.*), in the next year (Figure 1).

Figure 1. The periods used to partition the training and test sets and the periods used to compute the prediction target and the features for a patient and index year pair.



Data Set

We obtained a structured data set from the UWM enterprise data warehouse. This data set included administrative and clinical data relating to the patient cohort's encounters at the 3 hospitals and 12 clinics of the UWM from 2011 to 2020.

Features (Also Known as Independent Variables)

To improve model accuracy, we examined an extensive set of candidate features computed on the structured attributes in the data set. Table S1 of Multimedia Appendix 1 [3,18,28,30,50,59-83] shows these 278 candidate features coming from four sources: the known risk factors for COPD exacerbations [3,18,28,30,50,59-72], the features used in prior models to predict severe COPD exacerbations [20-53], the features that the clinician ZCL in our team suggested, and the features used in our prior models to predict asthma hospital encounters [84,85]. Asthma shares many similarities with COPD. Throughout this paper, whenever we mention the number of a given type of item (eg, medication) without using the word *distinct*, we count multiplicity.

Each input data instance to the predictive model contained 278 features, corresponded to a distinct patient and index year pair, and was used to predict the outcome of the patient in the next year. For this pair, the patient's age was computed based on the age at the end of the index year. The patient's primary care provider (PCP) was computed as the last recorded PCP of the patient by the end of the index year. The percentage of the PCP's patients with COPD in the preindex year having severe COPD exacerbations in the index year was computed on the data in the preindex and index years. Using the data from 2011 to the index year, we computed 26 features: the number of years from the first encounter related to COPD in the data set, the type of the first encounter related to COPD in the data set, 7 allergy features, and 17 features related to the problem list. The other 251 features were computed on the data in the index year.

Data Analysis

Data Preparation

Using the data preparation approach used in our papers [84,85], we identified the biologically implausible values, replaced them with null values, and normalized the data. As outcomes came from the next year, the data set had 9 years of effective data (2011-2019) over a time span of 10 years (2011-2020). To reflect future model use in clinical practice and to evaluate the impact of the COVID-19 pandemic on patient outcomes and model performance, we conducted two analyses:

1. Main analysis: we used the 2011-2018 data instances as the training set to train models and the 2019 data instances as the test set to assess model performance.
2. Performance stability analysis: we used the 2011-2017 data instances as the training set to train models and the 2018 data instances as the test set to assess model performance.

Classification Algorithms

We created machine learning classification models using Waikato Environment for Knowledge Analysis (WEKA; version 3.9) [86]. WEKA is a major open source software package for machine learning and data mining. It integrates many commonly used machine learning algorithms and feature selection techniques. We examined the 39 classification algorithms supported by WEKA and listed in the web-based multimedia appendix of our paper [84], as well as Extreme Gradient Boosting (XGBoost) [87] implemented in the XGBoost4J package [88]. XGBoost is a classification algorithm using an ensemble of decision trees. As XGBoost only takes numerical features, we converted categorical features to binary features through one-hot encoding. In the main analysis, we used the training set and our formerly published automatic machine learning model selection method [89] to automate the selection of the classification algorithm, feature selection technique, data balancing method to deal with imbalanced data, and hyperparameter values among all applicable ones. Compared

with the Auto-WEKA automatic machine learning model selection method [90], our method achieved an average of 11% (SD 15%) reduction in model error rate and a 28-fold reduction in search time. In the performance stability analysis, we used the same classification algorithm, feature selection technique, and hyperparameter values as those used in the final model of the main analysis.

Performance Metrics

As shown in the formulas, the performance of the models was evaluated with respect to the following metrics: accuracy (Table 1); sensitivity, also known as recall; specificity; positive predictive value (PPV), also known as precision; negative predictive value (NPV); and area under the receiver operating characteristic curve (AUC):

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad \textbf{(1)}$$

$$\text{Sensitivity} = TP / (TP + FN) \quad \textbf{(2)}$$

$$\text{Specificity} = TN / (TN + FP) \quad \textbf{(3)}$$

$$PPV = TP / (TP + FP) \quad \textbf{(4)}$$

$$NPV = TN / (TN + FN) \quad \textbf{(5)}$$

where TP stands for true positive, TN stands for true negative, FP stands for false positive, and FN stands for false negative.

Table 1. The confusion matrix.

Outcome class	Severe COPD[a] exacerbations in the next year	No severe COPD exacerbation in the next year
Predicted severe COPD exacerbations in the next year	True positive	False positive
Predicted no severe COPD exacerbation in the next year	False negative	True negative

[a]COPD: chronic obstructive pulmonary disease.

We computed the 95% CIs of the performance measures using the bootstrapping method [91]. We obtained 1000 bootstrap samples from the test set and computed the model's performance measures based on each bootstrap sample. This produced 1000 values for each performance metric. Their 2.5th and 97.5th percentiles provided the 95% CI of the corresponding performance measures. To depict the trade-off between sensitivity and specificity, we drew the receiver operating characteristic curve.

Results

Distributions of Data Instances and Bad Outcomes

The number of data instances increased over time. The proportion of data instances linked to bad outcomes remained relatively stable over time. The only exception was the sudden drop from 5.21% (369/7089) in 2018 to 2.42% (182/7529) in 2019 (Table 2), which resulted from the large drop in ED visits and inpatient stays for COPD in 2020 caused by the COVID-19 pandemic [92]. In the main analysis, 5.66% (2040/36,047) of the data instances in the training set and 2.42% (182/7529) of the data instances in the test set were linked to severe COPD exacerbations in the next year. In the performance stability analysis, 5.77% (1671/28,958) of the data instances in the training set and 5.21% (369/7089) of the data instances in the test set were linked to severe COPD exacerbations in the next year.

Table 2. The distributions of data instances and bad outcomes over time.

	Year								
	2011	2012	2013	2014	2015	2016	2017	2018	2019
Data instances, n	1848	2725	3204	4009	4875	5793	6504	7089	7529
Data instances linked to severe COPD[a] exacerbations in the next year, n (%)	128 (6.93)	176 (6.46)	183 (5.71)	223 (5.56)	272 (5.58)	351 (6.06)	338 (5.2)	369 (5.21)	182 (2.42)

[a]COPD: chronic obstructive pulmonary disease.

Patient Characteristics

Each patient and index year pair matched a data instance. For both the training set and the test set of the main analysis, when comparing the patient characteristic distributions between the data instances linked to severe COPD exacerbations in the next year and those linked to no severe COPD exacerbation in the next year, *P* values were computed using the chi-square 2-sample test and the Cochran–Armitage trend test [93] for categorical and numerical characteristics, respectively (Tables 3 and 4).

In the training set of the main analysis, most patient characteristics exhibited statistically significantly different distributions between the data instances linked to severe COPD exacerbations in the next year and those linked to no severe COPD exacerbation in the next year. Exceptions occurred on the patient characteristics of having prescriptions of inhaled corticosteroid, long-acting beta-2 agonist (LABA), and long-acting muscarinic antagonist (LAMA) combinations (*P*=.66); having prescriptions of phosphodiesterase-4 inhibitor (*P*=.06); presence of diabetes (*P*=.43); presence of eczema (*P*=.30); presence of lung cancer (*P*=.31); and presence of sinusitis (*P*=.61). In the test set of the main analysis, most patient characteristics exhibited statistically significantly different

distributions between the data instances linked to severe COPD exacerbations in the next year and those linked to no severe COPD exacerbation in the next year. Exceptions occurred on the patient characteristics of having private insurance (*P*=.79); having prescriptions of LABA and LAMA combinations (*P*=.54); having prescriptions of inhaled corticosteroid, LABA, and LAMA combinations (*P*=.90); having prescriptions of phosphodiesterase-4 inhibitor (*P*=.27); presence of allergic rhinitis (*P*=.24); presence of anxiety or depression (*P*=.08); presence of congestive heart failure (*P*=.11); presence of diabetes (*P*=.95); presence of eczema (*P*=.08); presence of hypertension (*P*=.05); presence of lung cancer (*P*=.51); presence of obesity (*P*=.25); presence of sinusitis (*P*=.99); and presence of sleep apnea (*P*=.22).

Table 3. The patient characteristics of the data instances in the training set of the main analysis.

Patient characteristic	Data instances (N=36,047), n (%)	Data instances linked to severe COPD[a] exacerbations in the next year (N=2040), n (%)	Data instances linked to no severe COPD exacerbation in the next year (N=34,007), n (%)	*P* value
Age (years)				*<.001* [b]
40-65	18,793 (52.13)	1219 (59.75)	17,574 (51.68)	*<.001*
>65	17,254 (47.87)	821 (40.25)	16,433 (48.32)	*<.001*
Sex				*<.001*
Female	15,414 (42.76)	749 (36.72)	14,665 (43.12)	*<.001*
Male	20,633 (57.24)	1291 (63.28)	19,342 (56.88)	*<.001*
Race				*<.001*
American Indian or Alaska Native	713 (1.98)	26 (1.27)	687 (2.02)	*<.001*
Asian	2092 (5.8)	144 (7.06)	1948 (5.73)	*<.001*
Black or African American	4795 (13.3)	524 (25.69)	4271 (12.56)	*<.001*
Native Hawaiian or other Pacific Islander	184 (0.51)	8 (0.39)	176 (0.52)	*<.001*
White	27,447 (76.14)	1330 (65.2)	26,117 (76.8)	*<.001*
Other, unknown, or not reported	816 (2.27)	8 (0.39)	808 (2.37)	*<.001*
Ethnicity				*<.001*
Hispanic	857 (2.38)	53 (2.6)	804 (2.36)	*<.001*
Non-Hispanic	32,585 (90.39)	1941 (95.15)	30,644 (90.11)	*<.001*
Unknown or not reported	2605 (7.23)	46 (2.25)	2559 (7.53)	*<.001*
Smoking status				*<.001*
Current smoker	16,952 (47.03)	1089 (53.38)	15,863 (46.65)	*<.001*
Former smoker	7367 (20.44)	345 (16.91)	7022 (20.65)	*<.001*
Never smoker or unknown	11,728 (32.53)	606 (29.71)	11,122 (32.7)	*<.001*
Insurance				
Private	17,513 (48.58)	834 (40.88)	16,679 (49.05)	*<.001*
Public	29,598 (82.11)	1767 (86.62)	27,831 (81.84)	*<.001*
Self-paid or charity	1994 (5.53)	229 (11.23)	1765 (5.19)	*<.001*
Number of years from the first encounter related to COPD in the data set				*<.001*
≤3	30,315 (84.1)	1566 (76.76)	28,749 (84.54)	*<.001*
>3	5732 (15.9)	474 (23.24)	5258 (15.46)	*<.001*
COPD medication prescription				
ICS[c]	13,327 (36.97)	1119 (54.85)	12,208 (35.9)	*<.001*
SAMA[d]	9608 (26.65)	1042 (51.08)	8566 (25.19)	*<.001*
SABA[e]	22,549 (62.55)	1684 (82.55)	20,865 (61.36)	*<.001*
SABA and SAMA combination	7174 (19.9)	810 (39.71)	6364 (18.71)	*<.001*
LAMA[f]	10,243 (28.42)	1001 (49.07)	9242 (27.18)	*<.001*
LABA[g]	8904 (24.7)	842 (41.27)	8062 (23.71)	*<.001*
LABA and LAMA combination	426 (1.18)	40 (1.96)	386 (1.14)	*.001*
ICS and LABA combination	8326 (23.1)	782 (38.33)	7544 (22.18)	*<.001*

Patient characteristic	Data instances (N=36,047), n (%)	Data instances linked to severe COPD[a] exacerbations in the next year (N=2040), n (%)	Data instances linked to no severe COPD exacerbation in the next year (N=34,007), n (%)	*P* value
ICS, LABA, and LAMA combination	16 (0.04)	0 (0)	16 (0.05)	.66
Phosphodiesterase-4 inhibitor	94 (0.26)	10 (0.49)	84 (0.25)	.06
Systemic corticosteroid	11,293 (31.33)	1144 (56.08)	10,149 (29.84)	*<.001*
Comorbidity				
Allergic rhinitis	2445 (6.78)	174 (8.53)	2271 (6.68)	*.001*
Anxiety or depression	10,786 (29.92)	725 (35.54)	10,061 (29.59)	*<.001*
Asthma	4794 (13.3)	417 (20.44)	4377 (12.87)	*<.001*
Congestive heart failure	6063 (16.82)	495 (24.26)	5568 (16.37)	*<.001*
Diabetes	7623 (21.15)	446 (21.86)	7177 (21.1)	.43
Eczema	1558 (4.32)	98 (4.8)	1460 (4.29)	.30
Gastroesophageal reflux	7162 (19.87)	507 (24.85)	6655 (19.57)	*<.001*
Hypertension	18,361 (50.94)	1150 (56.37)	17,211 (50.61)	*<.001*
Ischemic heart disease	7420 (20.58)	486 (23.82)	6934 (20.39)	*<.001*
Lung cancer	794 (2.2)	52 (2.55)	742 (2.18)	.31
Obesity	3487 (9.67)	255 (12.5)	3232 (9.5)	*<.001*
Sinusitis	1382 (3.83)	83 (4.07)	1299 (3.82)	.61
Sleep apnea	3179 (8.82)	253 (12.4)	2926 (8.6)	*<.001*

[a]COPD: chronic obstructive pulmonary disease.

[b]*P* value <.05 is italicized and signifies a statistically significant difference in the patient characteristic distributions.

[c]ICS: inhaled corticosteroid.

[d]SAMA: short-acting muscarinic antagonist.

[e]SABA: short-acting beta-2 agonist.

[f]LAMA: long-acting muscarinic antagonist.

[g]LABA: long-acting beta-2 agonist.

Table 4. The patient characteristics of the data instances in the test set of the main analysis.

Patient characteristic	Data instances (N=7529), n (%)	Data instances linked to severe COPD[a] exacerbations in the next year (N=182), n (%)	Data instances linked to no severe COPD exacerbation in the next year (N=7347), n (%)	*P* value
Age (years)				*<.001* [b]
40-65	3442 (45.72)	118 (64.8)	3324 (45.24)	*<.001*
>65	4087 (54.28)	64 (35.2)	4023 (54.76)	*<.001*
Sex				*<.001*
Female	3289 (43.68)	47 (25.8)	3242 (44.13)	*<.001*
Male	4240 (56.32)	135 (74.2)	4105 (55.87)	*<.001*
Race				*<.001*
American Indian or Alaska Native	156 (2.07)	5 (2.7)	151 (2.06)	*<.001*
Asian	439 (5.83)	7 (3.9)	432 (5.88)	*<.001*
Black or African American	896 (11.9)	57 (31.3)	839 (11.42)	*<.001*
Native Hawaiian or other Pacific Islander	53 (0.71)	2 (1.1)	51 (0.69)	*<.001*
White	5793 (76.94)	111 (61)	5682 (77.34)	*<.001*
Other, unknown, or not reported	192 (2.55)	0 (0)	192 (2.61)	*<.001*
Ethnicity				*.03*
Hispanic	188 (2.5)	3 (1.6)	185 (2.52)	*.03*
Non-Hispanic	7088 (94.14)	179 (98.4)	6909 (94.04)	*.03*
Unknown or not reported	253 (3.36)	0 (0)	253 (3.44)	*.03*
Smoking status				*.03*
Current smoker	3893 (51.71)	112 (61.5)	3781 (51.46)	*.03*
Former smoker	1267 (16.83)	25 (13.7)	1242 (16.91)	*.03*
Never smoker or unknown	2369 (31.47)	45 (24.7)	2324 (31.63)	*.03*
Insurance				
Private	4642 (61.65)	110 (60.4)	4532 (61.69)	.79
Public	6901 (91.66)	179 (98.4)	6722 (91.49)	*.002*
Self-paid or charity	540 (7.17)	41 (22.5)	499 (6.79)	*<.001*
Number of years from the first encounter related to COPD in the data set				*<.001*
≤3	5154 (68.46)	81 (44.5)	5073 (69.05)	*<.001*
>3	2375 (31.54)	101 (55.5)	2274 (30.95)	*<.001*
COPD medication prescription				
ICS[c]	2635 (35)	98 (53.8)	2537 (34.53)	*<.001*
SAMA[d]	1202 (15.96)	68 (37.4)	1134 (15.43)	*<.001*
SABA[e]	4241 (56.33)	158 (86.8)	4083 (55.57)	*<.001*
SABA and SAMA combination	1809 (24.03)	115 (63.2)	1694 (23.06)	*<.001*
LAMA[f]	2061 (27.37)	110 (60.4)	1951 (26.56)	*<.001*
LABA[g]	1760 (23.38)	77 (42.3)	1683 (22.91)	*<.001*
LABA and LAMA combination	400 (5.31)	12 (6.6)	388 (5.28)	.54

Patient characteristic	Data instances (N=7529), n (%)	Data instances linked to severe COPD[a] exacerbations in the next year (N=182), n (%)	Data instances linked to no severe COPD exacerbation in the next year (N=7347), n (%)	*P* value
ICS and LABA combination	1804 (23.96)	75 (41.2)	1729 (23.53)	*<.001*
ICS, LABA, and LAMA combination	69 (0.92)	1 (0.5)	68 (0.93)	.90
Phosphodiesterase-4 inhibitor	26 (0.35)	2 (1.1)	24 (0.33)	.27
Systemic corticosteroid	2385 (31.68)	103 (56.6)	2282 (31.06)	*<.001*
Comorbidity				
Allergic rhinitis	410 (5.45)	14 (7.7)	396 (5.39)	.24
Anxiety or depression	2153 (28.6)	63 (34.6)	2090 (28.45)	.08
Asthma	1096 (14.56)	43 (23.6)	1053 (14.33)	*<.001*
Congestive heart failure	1412 (18.75)	43 (23.6)	1369 (18.63)	.11
Diabetes	1689 (22.43)	40 (22)	1649 (22.44)	.95
Eczema	258 (3.43)	11 (6)	247 (3.36)	.08
Gastroesophageal reflux	1443 (19.17)	47 (25.8)	1396 (19)	*.03*
Hypertension	3791 (50.35)	105 (57.7)	3686 (50.17)	.05
Ischemic heart disease	1658 (22.02)	54 (29.7)	1604 (21.83)	*.02*
Lung cancer	203 (2.7)	3 (1.6)	200 (2.72)	.51
Obesity	669 (8.89)	21 (11.5)	648 (8.82)	.25
Sinusitis	279 (3.71)	7 (3.8)	272 (3.7)	.99
Sleep apnea	915 (12.15)	28 (15.4)	887 (12.07)	.22

[a]COPD: chronic obstructive pulmonary disease.

[b]*P* value <.05 is italicized and signifies a statistically significant difference in the patient characteristic distributions.

[c]ICS: inhaled corticosteroid.

[d]SAMA: short-acting muscarinic antagonist.

[e]SABA: short-acting beta-2 agonist.

[f]LAMA: long-acting muscarinic antagonist.

[g]LABA: long-acting beta-2 agonist.

Classification Algorithm and Features Used in the Final Model

The XGBoost algorithm was chosen by our automatic machine learning model selection method [89]. As a tree-based algorithm, XGBoost handles missing values in the features naturally. As detailed in Hastie et al [94], XGBoost automatically calculates an importance value for each feature based on the feature's apportioned contribution to the model. In the main analysis, the final model was created using XGBoost and the 229 features shown in descending order of their importance values in Table S2 of Multimedia Appendix 1. The other features contributed no extra predictive power and were automatically dropped by XGBoost.

Model Performance in the Main Analysis

In the main analysis with the test set, the final model had an AUC of 0.866 (95% CI 0.838-0.892), as computed from the model's receiver operating characteristic curve (Figure 2). The model's performance measures varied with the cutoff threshold for binary classification (Table 5). When using the top 9.99% (752/7529) of the patients with the largest predicted risk to set the cutoff threshold for binary classification, the model had an accuracy of 90.33% (6801/7529; 95% CI 89.61%-91.01%), a sensitivity of 56.6% (103/182; 95% CI 49.2%-64.2%), a specificity of 91.17% (6698/7347; 95% CI 90.51%-91.83%), a PPV of 13.7% (103/752; 95% CI 11.2%-16.2%), and an NPV of 98.83% (6698/6777; 95% CI 98.55%-99.08%), as computed from the corresponding confusion matrix of the model (Table 6).

Recall that 27 candidate features were computed on ≥2 years of data. When we ignored these features and considered only those computed with the data in the index year, the model's AUC dropped from 0.866 to 0.859 (95% CI 0.834-0.884). The top 19 features shown in Table S2 of Multimedia Appendix 1 have importance values ≥1%. When using only these features, the model's AUC dropped from 0.866 to 0.862 (95% CI 0.837-0.887). In this case, when using the top 9.99% (752/7529) of the patients with the largest predicted risk to set the cutoff threshold for binary classification, the model had an accuracy of 90.25% (6795/7529; 95% CI 89.56%-90.9%), a sensitivity of 54.9% (100/182; 95% CI 47.8%-61.9%), a specificity of

91.13% (6695/7347; 95% CI 90.43%-91.78%), a PPV of 13.3% (100/752; 95% CI 10.9%-15.7%), and an NPV of 98.79 (6695/6777; 95% CI 98.52%-99.06%).

Figure 2. The receiver operating characteristic curve of the final model in the main analysis.

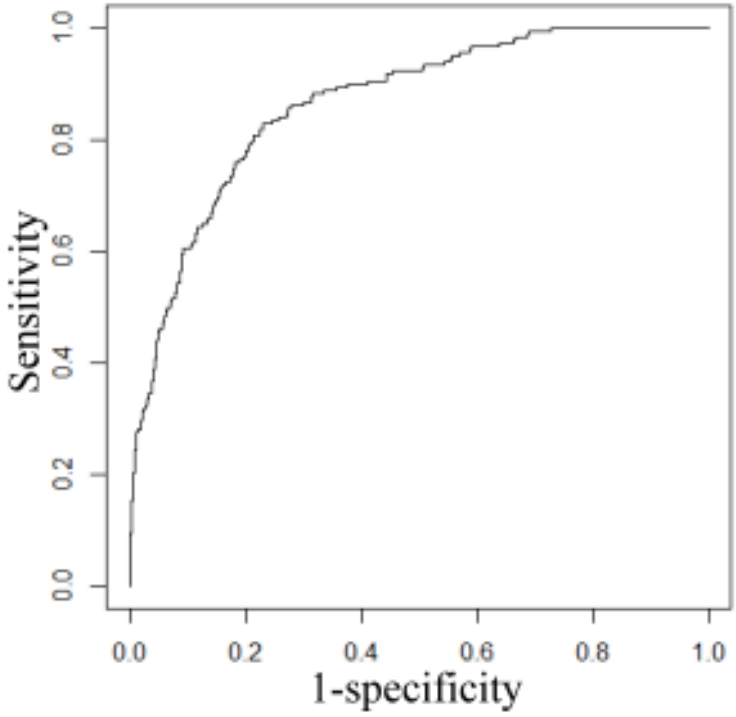


Table 5. In the main analysis, the performance measures of the final model with respect to using varying cutoff thresholds for binary classification.

Top percentage of patients with the largest predicted risk (%)	Accuracy (N=7529), n (%)	Sensitivity (N=182), n (%)	Specificity (N=7347), n (%)	Positive predictive value		Negative predictive value	
				n (%)	N	n (%)	N
1	7336 (97.4)	32 (17.6)	7304 (99.4)	32 (42.7)	75	7304 (98)	7454
2	7299 (96.9)	51 (28)	7248 (98.7)	51 (34)	150	7248 (98.2)	7379
3	7236 (96.1)	57 (31.3)	7179 (97.7)	57 (25.3)	225	7179 (98.3)	7304
4	7170 (95.2)	62 (34.1)	7108 (96.7)	62 (20.6)	301	7108 (98.3)	7228
5	7111 (94.4)	70 (38.5)	7041 (95.8)	70 (18.6)	376	7041 (98.4)	7153
6	7062 (93.8)	83 (45.6)	6979 (95)	83 (18.4)	451	6979 (98.6)	7078
7	6994 (92.9)	87 (47.8)	6907 (94)	87 (16.5)	527	6907 (98.6)	7002
8	6927 (92)	91 (50)	6836 (93)	91 (15.1)	602	6836 (98.7)	6927
9	6860 (91.1)	95 (52.2)	6765 (92.1)	95 (14)	677	6765 (98.7)	6852
10	6801 (90.3)	103 (56.6)	6698 (91.2)	103 (13.7)	752	6698 (98.8)	6777
15	6458 (85.8)	120 (65.9)	6338 (86.3)	120 (10.6)	1129	6338 (99)	6400
20	6118 (81.3)	138 (75.8)	5980 (81.4)	138 (9.2)	1505	5980 (99.3)	6024
25	5767 (76.6)	151 (83)	5616 (76.4)	151 (8)	1882	5616 (99.5)	5647

Table 6. The confusion matrix of the final model in the main analysis when using the top 9.99% (794/7944) of the patients with the largest predicted risk to set the cutoff threshold for binary classification.

Outcome class	Severe COPD[a] exacerbations in the next year	No severe COPD exacerbation in the next year
Predicted severe COPD exacerbations in the next year	103	649
Predicted no severe COPD exacerbation in the next year	79	6698

[a]COPD: chronic obstructive pulmonary disease.

Performance Stability Analysis

The final model in the main analysis and the model in the performance stability analysis had relatively similar performance (Table 7).

Table 7. The performance of the final model in the main analysis and the model in the performance stability analysis.

Performance measure	Final model in the main analysis[a]		Model in the performance stability analysis[b]	
	n (%; 95% CI)	N	n (%; 95% CI)	N
Accuracy	6801 (90.3; 89.6-91.0)	7529	6354 (89.6; 88.9-90.3)	7089
Sensitivity	103 (56.6; 49.2-64.2)	182	171 (46.3; 40.9-51.5)	369
Specificity	6698 (91.2; 90.5-91.8)	7347	6183 (92; 91.4-92.7)	6720
Positive predictive value	103 (13.7; 11.2-16.2)	752	171 (24.2; 20.8-27.2)	708
Negative predictive value	6698 (98.8; 98.6-99.1)	6777	6183 (96.9; 96.4-97.3)	6381

[a]Area under the receiver operating characteristic curve of 0.866 (95% CI 0.838-0.892).

[b]Area under the receiver operating characteristic curve of 0.847 (95% CI 0.828-0.864).

Discussion

Principal Findings

We created a machine learning model to predict severe COPD exacerbations in the next year in patients with COPD. The model had a higher AUC than the formerly published AUC of every prior model for predicting severe COPD exacerbations in the next year [20,25,27,28,30,33,35-43,46-49,51] (Tables 8 and 9). After improving our model's performance measures further (eg, by adding features extracted from clinical notes) and using our recently published automatic explanation method [95] to automatically explain the model's predictions, our model could be used as a decision support tool to advise the use of care management for patients with COPD and at high risk to improve outcomes.

In Table S2 of Multimedia Appendix 1, many of the top 19 features match the published (risk) factors that were highly correlated with COPD exacerbations, such as prior COPD exacerbations [18,60], prior health care encounters related to COPD [28,50], COPD medication use [50], BMI [70], peripheral capillary oxygen saturation [28], and heart rate [71].

We examined 278 candidate features, 82.4% (229/278) of which were used in the final model. Many omitted features are correlated with the outcome, but they provided no extra predictive power on the UWM data set beyond the 229 features used in the final model.

The prevalence rate of severe COPD exacerbations had a sudden drop in 2019. Despite this drop, our model still showed reasonably robust performance over time. This is desired for clinical decision support.

Table 8. A comparison of our final model and several prior models to predict severe chronic obstructive pulmonary disease (COPD) exacerbations in patients with COPD (Part 1).

Model	Data	Number of data instances	Prediction target (outcome)	Length of the period used to compute the outcome	Prevalence rate of the poor outcome (%)	Number of features checked	Classification algorithm	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	AUC[c]
Our final model	Administrative and clinical	43,576	ED[d] visit or inpatient stay for COPD	1 year	5.1	278	XGBoost[e]	56.6	91.17	13.7	98.83	0.866
Annavarapu et al [20]	Administrative	45,722	Inpatient stay for COPD	1 year	11.63	103	Logistic regression	17.3	97.5	48.1	90	0.77
Tavakoli et al [21]	Administrative	222,219	Inpatient stay for COPD	2 months	1.02	83	Gradient boosting	23	98	—[f]	—	0.820
Samp et al [22]	Administrative	478,772	Inpatient stay for COPD	6 months	2.2	101	Logistic regression	17.6	96.6	—	—	—
Thomsen et al [23]	Research	6574	Two or more exacerbations (medication change or inpatient stay for COPD)	1-7 years	6.4	11	Logistic regression	—	—	18	96	0.73
Orchard et al [24]	Research	57,150	Inpatient stay for COPD	1 day	0.1	153	Neural network	80	60	—	—	0.740
Suetomo et al [25]	Research	123	Inpatient stay for COPD	1 year	12.2	18	Logistic regression	53	49	—	—	0.79
Lee et al [26]	Research and clinical	545	Medication change, ED visit, or inpatient stay for COPD	6 months	46	10	Logistic regression	52	69	—	—	0.63
Faganello et al [27]	Research	120	Outpatient, inpatient, or ED encounter for COPD	1 year	50	16	Logistic regression	58.3	73.3	—	—	0.686
Alcázar et al [28]	Research	127	Inpatient stay for COPD	1 year	39.4	9	Logistic regression	76.2	77.3	61.5	87.2	0.809
Bertens et al [29]	Research and clinical	1033	Medication change or inpatient stay for COPD	2 years	28.3	7	Logistic regression	—	—	—	—	0.66
Miravitlles et al [30]	Research and clinical	713	Inpatient stay for COPD	1 year	22.2	7	Logistic regression	—	—	—	—	0.582
Make et al [31]	Research	3141	Medication change, ED visit, or inpatient stay for COPD	6 months	—	38	Logistic regression	—	—	—	—	0.67
Montserrat-Capdevila et al [32]	Administrative and clinical	2501	Inpatient stay for COPD	3 years	32.5	17	Logistic regression	—	—	—	—	0.72

Model	Data	Number of data instances	Prediction target (outcome)	Length of the period used to compute the outcome	Prevalence rate of the poor outcome (%)	Number of features checked	Classification algorithm	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	AUC[c]
Kerkhof et al [33]	Research and clinical	16,565	Two or more exacerbations (medication change, ED visit, or inpatient stay for COPD)	1 year	19.6	22	Logistic regression	—	—	—	—	0.735
Chen et al [34]	Research	1711	ED visit or inpatient stay for COPD	5 years	30.6	14	Cox proportional hazard regression	—	—	—	—	0.74
Yii et al [35]	Administrative and clinical	237	Inpatient stay for COPD	1 year	1.41 per patient year	31	Negative binomial regression	—	—	—	—	0.789

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]AUC: area under the receiver operating characteristic curve.

[d]ED: emergency department.

[e]XGBoost: Extreme Gradient Boosting.

[f]The performance measure is unreported in the initial paper describing the model.

Table 9. A comparison of our final model and several prior models to predict severe chronic obstructive pulmonary disease (COPD) exacerbations in patients with COPD (Part 2).

Model	Data	Number of data instances	Prediction target (outcome)	Length of the period used to compute the outcome	Prevalence rate of the poor outcome (%)	Number of features checked	Classification algorithm	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	AUC[c]
Our final model	Administrative and clinical	43,576	ED[d] visit or inpatient stay for COPD	1 year	5.1	278	XGBoost[e]	56.6	91.17	13.7	98.83	0.866
Adibi et al [36]	Research	2380	ED visit or inpatient stay for COPD	1 year	0.29 per year	13	Mixed effect logistic	—[f]	—	—	—	0.77
Stanford et al [37]	Administrative	258,668	Inpatient stay for COPD	1 year	8.5	30	Logistic regression	—	—	—	—	0.749
Stanford et al [38]	Administrative	223,824	Inpatient stay for COPD	1 year	6.63	30	Logistic regression	—	—	—	—	0.711
Stanford et al [39]	Administrative	92,496	Inpatient stay for COPD	1 year	—	30	Logistic regression	—	—	—	—	0.801
Stanford et al [40]	Administrative	60,776	Inpatient stay for COPD	1 year	19.16	8	Logistic regression	—	—	—	—	0.742
Jones et al [41]	Clinical	375	Inpatient stay for COPD	1 year	—	4	Index	—	—	—	—	0.755
Jones et al [42]	Research and clinical	7105	Inpatient stay for COPD	1 year	—	8	Negative binomial regression	—	—	—	—	0.64
Fan et al [43]	Research	3282	Inpatient stay for COPD	1 year	4.3	23	Logistic regression	—	—	—	—	0.706
Moy et al [44]	Research and clinical	167	Inpatient stay for COPD	4-21 months	32.9	6	Negative binomial regression	—	—	—	—	0.69
Briggs et al [45]	Research	8802	Inpatient stay for COPD	6 months to 3 years	9	13	Cox proportional hazard regression	—	—	—	—	0.71
Lange et al [46]	Administrative and research	6628	Medication change or inpatient stay for COPD	1 year	4.8	3	GOLD[g] stratification	—	—	—	—	0.7
Abascal-Bolado et al [47]	Research and clinical	493	Inpatient stay for COPD	1 year	—	8	Classification and regression tree	—	—	—	—	0.70
Blanco-Aparicio et al [48]	Research	100	ED visit for COPD	1 year	21	12	Logistic regression	—	—	—	—	0.651
Yoo et al [49]	Research and clinical	260	Medication change, ED visit, or inpatient stay for COPD	1 year	40.8	17	Logistic regression	—	—	—	—	0.69

Model	Data	Number of data instances	Prediction target (outcome)	Length of the period used to compute the outcome	Prevalence rate of the poor outcome (%)	Number of features checked	Classification algorithm	Sensitivity (%)	Specificity (%)	PPV[a] (%)	NPV[b] (%)	AUC[c]
Niewoehner et al [50]	Research and clinical	1829	Inpatient stay for COPD	6 months	8.3	27	Cox proportional hazard regression	—	—	—	—	0.73
Austin et al [51]	Administrative	638,926	COPD-related inpatient stay	1 year	—	34	Logistic regression	—	—	—	—	0.778
Marin et al [52]	Research	275	Inpatient stay for COPD	6 months to 8 years	—	4	Logistic regression	86	73	—	—	0.88
Marin et al [52]	Research	275	ED visit for COPD	6 months to 8 years	—	4	Logistic regression	58	87	—	—	0.78
Ställberg et al [53]	Administrative and clinical	7823	COPD-related inpatient stay	10 days	—	>4000	XGBoost	16	—	11	—	0.86

[a]PPV: positive predictive value.

[b]NPV: negative predictive value.

[c]AUC: area under the receiver operating characteristic curve.

[d]ED: emergency department.

[e]XGBoost: Extreme Gradient Boosting.

[f]The performance measure is unreported in the initial paper describing the model.

[g]GOLD: Global Initiative for Chronic Obstructive Lung Disease.

Comparison With Prior Work

Researchers formerly created several models to predict severe COPD exacerbations in patients with COPD [20-53]. Tables 8 and 9 present comparisons between our final model and these models, which include all related models listed in the systematic reviews by Guerra et al [96] and Bellou et al [97] as well as several recent models that were published after the reviews. Our final model predicted severe COPD exacerbations in the next year. Every prior model for predicting severe COPD exacerbations in the next year had an AUC ≤0.809, that is, at least 0.057 lower than that of our final model. Compared with the prior models for predicting severe COPD exacerbations other than the model developed by Ställberg et al [53], our final model used more extensive features with predictive power, which helped improve model performance.

Our final model's prediction target covered both future ED visits and future inpatient stays for COPD, which we want to use care management to prevent. Among all prior models, only 2 [34,36] had prediction targets covering both future ED visits and future inpatient stays for COPD. Most of the prior models predicted either only future ED visits [48,52] or only future inpatient stays for COPD [20-22,24,25,28,30,32,35,37-45,47,50-52]. This would be insufficient for preventing both future ED visits and future inpatient stays for COPD. The other prior models [23,26,27,29,31,33,46,49] had prediction targets covering both moderate and severe COPD exacerbations, with moderate COPD exacerbations typically referring to COPD medication change such as the use of systemic corticosteroids. These prediction targets were not specific enough for identifying patients at the highest risk for care management because a care management program can host only a small portion of patients [17].

To make it suitable for use in daily clinical practice, our final model was built on routinely available administrative and clinical data. In comparison, the models developed by several other research groups [23-31,33,34,36,42-50,52] used research data, some of which are unavailable in usual clinical practice. Thus, these models would be unsuitable for daily clinical use.

Our predictive model was developed to guide COPD care management's enrollment decisions and to prevent severe COPD exacerbations. To give enough lead time for preventive interventions to be effective and to use precious care management resources well, we chose severe COPD exacerbation in the next year as the prediction target. In comparison, the model developed by Orchard et al [24] predicted inpatient stays for COPD on the next day. If a patient will incur an inpatient stay for COPD tomorrow, intervening starting from today could be too late to avoid the inpatient stay. At present, we are aware of no published conclusion on how long it will take for any intervention to be effective at preventing severe COPD exacerbations. In the studies by Longman et al [98] and Johnston et al [99], several clinicians had expressed the opinion that it could take as long as 3 months for any intervention to be

effective at preventing inpatient stays for a chronic, ambulatory care–sensitive condition. Our final model will have a different clinical use from the models that make short-term predictions. Foreseeing a severe COPD exacerbation in the next 12 months would be useful for identifying and personalizing medium-term interventions and maintenance therapies to change the course of the disease. In comparison, foreseeing a severe COPD exacerbation in the next 1 or few days can be useful for deciding acute management approaches to improve outcomes, such as preemptive hospitalization of the patient to avoid more severe adverse outcomes, but would be inadequate for trying to improve the course of the disease in a short amount of time. In fact, treatment approaches proven to be effective at reducing severe COPD exacerbations are usually not indicated for acute management.

Marin et al [52] built a model to predict inpatient stays for COPD in up to the next 8 years with an AUC of 0.88 and a separate model to predict ED visits for COPD in up to the next 8 years with an AUC of 0.78. An inpatient stay or an ED visit that will happen several years later is too remote to be worth using precious care management resources now to prevent.

For the patients with COPD who will have severe COPD exacerbations in the future, sensitivity is the proportion of patients whom the model identifies. The difference in sensitivity could greatly affect hospital use. Our final model's sensitivity is higher than the sensitivities achieved by the models developed by several other research groups [20-22,25,26,53]. Compared with our final model, the models developed by Orchard et al [24], Faganello et al [27], and Alcázar et al [28] each reached a higher sensitivity at the price of a much lower specificity. For each of these 3 models, if we adjust the cutoff threshold for binary classification and make our final model have the same specificity as that model, our final model would achieve a higher sensitivity than that model. More specifically, at a specificity of 60.02% (4410/7347), our final model achieved a sensitivity of 90.1% (164/182), whereas the model developed by Orchard et al [24] achieved a sensitivity of 80%. At a specificity of 73.3% (5385/7347), our final model achieved a sensitivity of 84.1% (153/182), whereas the model developed by Faganello et al [27] achieved a sensitivity of 58.3%. At a specificity of 77.34% (5682/7347), our final model achieved a sensitivity of 81.9% (149/182), whereas the model developed by Alcázar et al [28] achieved a sensitivity of 76.2%.

The prevalence rate of poor outcomes has a large impact on any model's PPV [100]. On our data set, where this prevalence rate is approximately 5%, our final model reached a PPV of <14%. In comparison, on a data set where this prevalence rate is 11.63%, the model developed by Annavarapu et al [20] reached a PPV of 48.1%. On a data set where this prevalence rate is 6.4%, the model developed by Thomsen et al [23] reached a PPV of 18%. On a data set where this prevalence rate is 39.4%, the model developed by Alcázar et al [28] reached a PPV of 61.5%. In all 3 cases, the higher prevalence rates of poor outcomes permitted the PPV to be larger.

Our data set is imbalanced, with only a small portion of patients to have severe COPD exacerbations in the next year. For imbalanced data sets, the area under the precision–recall curve (AUPRC) is a better measure of overall model performance than the AUC [101]. The AUPRC was reported for only the model developed by Ställberg et al [53] among all the prior models. Although the model developed by Ställberg et al [53] had an AUC of 0.86, which is only slightly lower than that of our final model, our final model had an AUPRC of 0.24 (95% CI 0.18-0.31) that is 3 times as large as the 0.08 AUPRC of that model. In addition, that model predicted COPD-related inpatient stays, for which COPD can be any of the diagnoses, in the next 10 days. If a patient will incur an inpatient stay in the next 10 days, intervening starting from today could be too late to avoid the inpatient stay. In comparison, our final model predicted ED visits or inpatient stays with a principal diagnosis of COPD in the next year, allowing more lead time for preventive interventions to be effective.

Considerations for Future Clinical Use

Our final model reached an AUC that is larger than every AUC formerly reported in the literature for predicting severe COPD exacerbations in the next year. Despite having a relatively low PPV, our final model could still benefit health care for 3 reasons.

First, health care systems such as the UWM and Intermountain Healthcare use proprietary models, which have similar performance to the formerly published models, to allocate COPD care management resources. Our final model had a higher AUC than all formerly reported AUCs for predicting severe COPD exacerbations in the next year. Hence, although we plan to investigate using various techniques to further improve model performance in the future, we think it is already worth considering using our final model to replace the proprietary models currently being used at health care systems such as the UWM for COPD care management.

Second, we set the cutoff threshold for binary classification at the top 9.99% (752/7529) of the patients with the largest predicted risk. In this case, a perfect model would achieve the theoretically maximum possible PPV of 24.2% (182/752). Our final model's PPV is 56.6% (103/182) of the theoretically maximum possible PPV. In other words, our final model captured 56.6% (103/182) of the patients with COPD who would have severe COPD exacerbations in the next year. If we change the cutoff threshold to the top 25% of the patients with the largest predicted risk, the final model would capture 83% (151/182) of the patients with COPD who would have severe COPD exacerbations in the next year.

Third, a PPV at the level of our final model's PPV is suitable for identifying patients with COPD and at high risk for low-cost preventive interventions such as arranging a nurse to further follow up with the patient through phone calls, teaching the patient to correctly use a COPD inhaler, teaching the patient the correct use of a peak flow meter to self-monitor symptoms at home, and enrolling the patient in a home-based pulmonary rehabilitation program [102].

Our final model used 229 features. To ease clinical deployment, we could reduce features, for example, to the top 19 with importance values ≥1%. A feature's importance value differs across health care systems. If conditions permit, we should use

a data set from the target health care system to compute the features' importance values and decide which features to retain.

Our final model was based on XGBoost [87], which leverages the hyperparameter scale_pos_weight to balance the weights of the 2 outcome classes in our data set [103]. The scale_pos_weight hyperparameter was set by our automatic model selection method [89] to a nondefault value to maximize our final model's AUC [104]. This caused the side effect of greatly increasing our model's predicted probabilities of having future severe COPD exacerbations to values much larger than the true probabilities [103]. However, it does not affect our ability to identify the top portion of the patients with the largest predicted risk for preventive interventions. If preferred, we could forgo the balancing by keeping scale_pos_weight at its default value 1. In this case, our model's AUC would drop by 0.003 to 0.863 (95% CI 0.835-0.888), which is still larger than every formerly published AUC for predicting severe COPD exacerbations in the next year.

Limitations

This study includes several limitations that are worth future work.

First, this study used solely structured data. It is worth considering performing natural language processing to extract features from unstructured clinical notes to improve model performance. A model with higher performance can be used to better facilitate COPD care management.

Second, this study used age, diagnosis codes, and medication data to identify patients with COPD and used diagnosis codes and encounter information to define the prediction target. One can use age, diagnosis codes, and medication data to identify patients with COPD reasonably well [56]; yet, diagnosis codes were shown to have a low sensitivity in capturing inpatient stays for COPD [105]. Our predictive model is likely to perform poorly at finding those patients who would experience only future inpatient stays for COPD that are not captured by our current definition of the prediction target. We expect that this will not greatly affect our predictive model's usefulness for facilitating COPD care management. On the basis of our current definition of the prediction target, >5% of the patients in our data set had severe COPD exacerbations in the following year. If fully captured by the predictive model, these patients would have already exceeded the service capacity of a typical care management program, which can take ≤3% of the patients [17]. In the future, one could consider adding both medication data and information extracted from clinical notes through natural language processing to better capture inpatient stays for COPD.

Third, this study used non–deep learning classification algorithms. Deep learning has improved model performance for many clinical predictive modeling tasks [106-111]. It is worth investigating whether using deep learning can improve model performance for predicting severe COPD exacerbations.

Fourth, this study used data from a single health care system: the UWM. It is worth evaluating our model's generalizability to other health care systems. We are working on obtaining a data set of patients with COPD from Intermountain Healthcare for this purpose [112].

Fifth, our data set contained no information on UWM patients' health care use at other health care systems. It is worth evaluating how our model's performance would change if data on UWM patients' health care use at other health care systems are available.

Conclusions

This work improved the state of the art of predicting severe COPD exacerbations in patients with COPD. In particular, our final model had a higher AUC than every formerly published model AUC on predicting severe COPD exacerbations in the next year. After improving our model's performance measures further and using our recently published automatic explanation method [95] to automatically explain the model's predictions, our model could be used in a decision support tool to guide the use of care management for patients with COPD and at high risk to improve outcomes.

Acknowledgments

GL and SZ were partially supported by the National Heart, Lung, and Blood Institute of the National Institutes of Health under award number R01HL142503. SZ was also partially supported by the National Library of Medicine Training Grant under award number T15LM007442. MA was partially supported by grants from the Flight Attendant Medical Research Institute (CIA190001) and the California Tobacco-Related Disease Research Program (T29IR0715). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. YT did the work at the University of Washington when she was a visiting PhD student.

Authors' Contributions

GL and SZ were mainly responsible for the paper. SZ performed a literature review, extracted and analyzed the data, constructed the models, and wrote the first draft of the paper. GL conceptualized and designed the study, participated in performing data analysis, and rewrote the whole paper. MA and ZCL provided clinical expertise, contributed to conceptualizing the presentation, and revised the paper. YT took part in extracting the data and identifying the biologically implausible values.

Conflicts of Interest

None declared.

Multimedia Appendix 1
The candidate features and the features used in the final model in the main analysis and their importance values.
[PDF File (Adobe PDF File), 190 KB - jmir_v24i1e28953_app1.pdf]

References

1. Ford ES, Murphy LB, Khavjou O, Giles WH, Holt JB, Croft JB. Total and state-specific medical and absenteeism costs of COPD among adults aged ≥ 18 years in the United States for 2010 and projections through 2020. Chest 2015 Jan;147(1):31-45. [doi: 10.1378/chest.14-0972] [Medline: 25058738]
2. Disease or condition of the week - COPD. Centers for Disease Control and Prevention. 2019. URL: https://www.cdc.gov/dotw/copd/index.html [accessed 2021-12-20]
3. 2020 Gold reports. Global Initiative for Chronic Obstructive Lung Disease - GOLD. 2020. URL: https://goldcopd.org/gold-reports [accessed 2021-12-20]
4. Blanchette CM, Dalal AA, Mapel D. Changes in COPD demographics and costs over 20 years. J Med Econ 2012;15(6):1176-1182. [doi: 10.3111/13696998.2012.713880] [Medline: 22812689]
5. Anzueto A, Leimer I, Kesten S. Impact of frequency of COPD exacerbations on pulmonary function, health status and clinical outcomes. Int J Chron Obstruct Pulmon Dis 2009;4:245-251 [FREE Full text] [doi: 10.2147/copd.s4862] [Medline: 19657398]
6. Connors Jr AF, Dawson NV, Thomas C, Harrell Jr FE, Desbiens N, Fulkerson WJ, et al. Outcomes following acute exacerbation of severe chronic obstructive lung disease. The SUPPORT investigators (Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatments). Am J Respir Crit Care Med 1996 Oct;154(4 Pt 1):959-967. [doi: 10.1164/ajrccm.154.4.8887592] [Medline: 8887592]
7. Viglio S, Iadarola P, Lupi A, Trisolini R, Tinelli C, Balbi B, et al. MEKC of desmosine and isodesmosine in urine of chronic destructive lung disease patients. Eur Respir J 2000 Jun;15(6):1039-1045 [FREE Full text] [doi: 10.1034/j.1399-3003.2000.01511.x] [Medline: 10885422]
8. Kanner RE, Anthonisen NR, Connett JE, Lung Health Study Research Group. Lower respiratory illnesses promote FEV(1) decline in current smokers but not ex-smokers with mild chronic obstructive pulmonary disease: results from the lung health study. Am J Respir Crit Care Med 2001 Aug 01;164(3):358-364. [doi: 10.1164/ajrccm.164.3.2010017] [Medline: 11500333]
9. Spencer S, Jones PW, GLOBE Study Group. Time course of recovery of health status following an infective exacerbation of chronic bronchitis. Thorax 2003 Jul;58(7):589-593 [FREE Full text] [doi: 10.1136/thorax.58.7.589] [Medline: 12832673]
10. Spencer S, Calverley PM, Burge PS, Jones PW, ISOLDE Study Group. Inhaled Steroids in Obstructive Lung Disease. Health status deterioration in patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2001 Jan;163(1):122-128. [doi: 10.1164/ajrccm.163.1.2005009] [Medline: 11208636]
11. Johnston J, Longman J, Ewald D, King J, Das S, Passey M. Study of potentially preventable hospitalisations (PPH) for chronic conditions: what proportion are preventable and what factors are associated with preventable PPH? BMJ Open 2020 Nov 09;10(11):e038415 [FREE Full text] [doi: 10.1136/bmjopen-2020-038415] [Medline: 33168551]
12. Billings J, Zeitel L, Lukomnik J, Carey TS, Blank AE, Newman L. Impact of socioeconomic status on hospital use in New York City. Health Aff (Millwood) 1993;12(1):162-173. [doi: 10.1377/hlthaff.12.1.162] [Medline: 8509018]
13. Mays GP, Claxton G, White J. Managed care rebound? Recent changes in health plans' cost containment strategies. Health Aff (Millwood) 2004;Suppl Web Exclusives:427-436 [FREE Full text] [doi: 10.1377/hlthaff.w4.427] [Medline: 15451964]
14. Rice KL, Dewan N, Bloomfield HE, Grill J, Schult TM, Nelson DB, et al. Disease management program for chronic obstructive pulmonary disease: a randomized controlled trial. Am J Respir Crit Care Med 2010 Oct 1;182(7):890-896. [doi: 10.1164/rccm.200910-1579OC] [Medline: 20075385]
15. Bandurska E, Damps-Konstańska I, Popowski P, Jędrzejczyk T, Janowiak P, Świętnicka K, et al. Impact of integrated care model (ICM) on direct medical costs in management of advanced chronic obstructive pulmonary disease (COPD). Med Sci Monit 2017 Jun 12;23:2850-2862 [FREE Full text] [doi: 10.12659/msm.901982] [Medline: 28603270]
16. Curry N, Billings J, Darin B, Dixon J, Williams M, Wennberg D. Predictive risk project literature review. King's Fund, London. 2005. URL: http://www.kingsfund.org.uk/sites/files/kf/field/field_document/predictive-risk-literature-review-june2005.pdf, [accessed 2021-12-20]
17. Axelrod RC, Vogel D. Predictive modeling in health plans. Dis Manag Health Outcomes 2003;11(12):779-787. [doi: 10.2165/00115677-200311120-00003]
18. Hurst JR, Vestbo J, Anzueto A, Locantore N, Müllerova H, Tal-Singer R, Evaluation of COPD Longitudinally to Identify Predictive Surrogate Endpoints (ECLIPSE) Investigators. Susceptibility to exacerbation in chronic obstructive pulmonary disease. N Engl J Med 2010 Sep 16;363(12):1128-1138. [doi: 10.1056/NEJMoa0909883] [Medline: 20843247]
19. Blagev DP, Collingridge DS, Rea S, Press VG, Churpek MM, Carey K, et al. Stability of frequency of severe chronic obstructive pulmonary disease exacerbations and health care utilization in clinical populations. Chronic Obstr Pulm Dis 2018 Jun 20;5(3):208-220 [FREE Full text] [doi: 10.15326/jcopdf.5.3.2017.0183] [Medline: 30584584]
20. Annavarapu S, Goldfarb S, Gelb M, Moretz C, Renda A, Kaila S. Development and validation of a predictive model to identify patients at risk of severe COPD exacerbations using administrative claims data. Int J Chron Obstruct Pulmon Dis 2018;13:2121-2130 [FREE Full text] [doi: 10.2147/COPD.S155773] [Medline: 30022818]

21. Tavakoli H, Chen W, Sin DD, FitzGerald JM, Sadatsafavi M. Predicting severe chronic obstructive pulmonary disease exacerbations. Developing a population surveillance approach with administrative data. Ann Am Thorac Soc 2020 Sep;17(9):1069-1076. [doi: 10.1513/AnnalsATS.202001-070OC] [Medline: 32383971]
22. Samp JC, Joo MJ, Schumock GT, Calip GS, Pickard AS, Lee TA. Predicting acute exacerbations in chronic obstructive pulmonary disease. J Manag Care Spec Pharm 2018 Mar;24(3):265-279. [doi: 10.18553/jmcp.2018.24.3.265] [Medline: 29485951]
23. Thomsen M, Ingebrigtsen TS, Marott JL, Dahl M, Lange P, Vestbo J, et al. Inflammatory biomarkers and exacerbations in chronic obstructive pulmonary disease. J Am Med Assoc 2013 Jun 12;309(22):2353-2361. [doi: 10.1001/jama.2013.5732] [Medline: 23757083]
24. Orchard P, Agakova A, Pinnock H, Burton CD, Sarran C, Agakov F, et al. Improving prediction of risk of hospital admission in chronic obstructive pulmonary disease: application of machine learning to telemonitoring data. J Med Internet Res 2018 Sep 21;20(9):e263 [FREE Full text] [doi: 10.2196/jmir.9227] [Medline: 30249589]
25. Suetomo M, Kawayama T, Kinoshita T, Takenaka S, Matsuoka M, Matsunaga K, et al. COPD assessment tests scores are associated with exacerbated chronic obstructive pulmonary disease in Japanese patients. Respir Investig 2014 Sep;52(5):288-295. [doi: 10.1016/j.resinv.2014.04.004] [Medline: 25169844]
26. Lee SD, Huang MS, Kang J, Lin CH, Park MJ, Oh YM, Investigators of the Predictive Ability of CAT in Acute Exacerbations of COPD (PACE) Study. The COPD assessment test (CAT) assists prediction of COPD exacerbations in high-risk patients. Respir Med 2014 Apr;108(4):600-608 [FREE Full text] [doi: 10.1016/j.rmed.2013.12.014] [Medline: 24456695]
27. Faganello MM, Tanni SE, Sanchez FF, Pelegrino NR, Lucheta PA, Godoy I. BODE index and GOLD staging as predictors of 1-year exacerbation risk in chronic obstructive pulmonary disease. Am J Med Sci 2010 Jan;339(1):10-14. [doi: 10.1097/MAJ.0b013e3181bb8111] [Medline: 19926966]
28. Alcázar B, García-Polo C, Herrejón A, Ruiz LA, de Miguel J, Ros JA, et al. Factors associated with hospital admission for exacerbation of chronic obstructive pulmonary disease. Arch Bronconeumol 2012 Mar;48(3):70-76. [doi: 10.1016/j.arbres.2011.10.009] [Medline: 22196478]
29. Bertens LC, Reitsma JB, Moons KG, van Mourik Y, Lammers JW, Broekhuizen BD, et al. Development and validation of a model to predict the risk of exacerbations in chronic obstructive pulmonary disease. Int J Chron Obstruct Pulmon Dis 2013;8:493-499 [FREE Full text] [doi: 10.2147/COPD.S49609] [Medline: 24143086]
30. Miravitlles M, Guerrero T, Mayordomo C, Sánchez-Agudo L, Nicolau F, Segú JL. Factors associated with increased risk of exacerbation and hospital admission in a cohort of ambulatory COPD patients: a multiple logistic regression analysis. The EOLO Study Group. Respiration 2000;67(5):495-501. [doi: 10.1159/000067462] [Medline: 11070451]
31. Make BJ, Eriksson G, Calverley PM, Jenkins CR, Postma DS, Peterson S, et al. A score to predict short-term risk of COPD exacerbations (SCOPEX). Int J Chron Obstruct Pulmon Dis 2015;10:201-209 [FREE Full text] [doi: 10.2147/COPD.S69589] [Medline: 25670896]
32. Montserrat-Capdevila J, Godoy P, Marsal JR, Barbé F. Predictive model of hospital admission for COPD exacerbation. Respir Care 2015 Sep;60(9):1288-1294 [FREE Full text] [doi: 10.4187/respcare.04005] [Medline: 26286737]
33. Kerkhof M, Freeman D, Jones R, Chisholm A, Price DB, Respiratory Effectiveness Group. Predicting frequent COPD exacerbations using primary care data. Int J Chron Obstruct Pulmon Dis 2015;10:2439-2450 [FREE Full text] [doi: 10.2147/COPD.S94259] [Medline: 26609229]
34. Chen X, Wang Q, Hu Y, Zhang L, Xiong W, Xu Y, et al. A nomogram for predicting severe exacerbations in stable COPD patients. Int J Chron Obstruct Pulmon Dis 2020;15:379-388 [FREE Full text] [doi: 10.2147/COPD.S234241] [Medline: 32110006]
35. Yii AC, Loh CH, Tiew PY, Xu H, Taha AA, Koh J, et al. A clinical prediction model for hospitalized COPD exacerbations based on "treatable traits". Int J Chron Obstruct Pulmon Dis 2019;14:719-728 [FREE Full text] [doi: 10.2147/COPD.S194922] [Medline: 30988606]
36. Adibi A, Sin DD, Safari A, Johnson KM, Aaron SD, FitzGerald JM, et al. The Acute COPD Exacerbation Prediction Tool (ACCEPT): a modelling study. Lancet Respir Med 2020 Oct;8(10):1013-1021. [doi: 10.1016/S2213-2600(19)30397-2] [Medline: 32178776]
37. Stanford RH, Nag A, Mapel DW, Lee TA, Rosiello R, Vekeman F, et al. Validation of a new risk measure for chronic obstructive pulmonary disease exacerbation using health insurance claims data. Ann Am Thorac Soc 2016 Jul;13(7):1067-1075. [doi: 10.1513/AnnalsATS.201508-493OC] [Medline: 27070274]
38. Stanford RH, Nag A, Mapel DW, Lee TA, Rosiello R, Schatz M, et al. Claims-based risk model for first severe COPD exacerbation. Am J Manag Care 2018 Feb 1;24(2):45-53 [FREE Full text] [Medline: 29461849]
39. Stanford RH, Lau MS, Li Y, Stemkowski S. External validation of a COPD risk measure in a commercial and Medicare population: the COPD treatment ratio. J Manag Care Spec Pharm 2019 Jan;25(1):58-69. [doi: 10.18553/jmcp.2019.25.1.058] [Medline: 30589629]
40. Stanford RH, Korrer S, Brekke L, Reinsch T, Bengtson LG. Validation and assessment of the COPD treatment ratio as a predictor of severe exacerbations. Chronic Obstr Pulm Dis 2020 Jan;7(1):38-48 [FREE Full text] [doi: 10.15326/jcopdf.7.1.2019.0132] [Medline: 31999901]

41. Jones RC, Donaldson GC, Chavannes NH, Kida K, Dickson-Spillmann M, Harding S, et al. Derivation and validation of a composite index of severity in chronic obstructive pulmonary disease: the DOSE Index. Am J Respir Crit Care Med 2009 Dec 15;180(12):1189-1195. [doi: 10.1164/rccm.200902-0271OC] [Medline: 19797160]
42. Jones RC, Price D, Chavannes NH, Lee AJ, Hyland ME, Ställberg B, UNLOCK Group of the IPCRG. Multi-component assessment of chronic obstructive pulmonary disease: an evaluation of the ADO and DOSE indices and the global obstructive lung disease categories in international primary care data sets. NPJ Prim Care Respir Med 2016 Apr 07;26:16010 [FREE Full text] [doi: 10.1038/npjpcrm.2016.10] [Medline: 27053297]
43. Fan VS, Curtis JR, Tu SP, McDonell MB, Fihn SD, Ambulatory Care Quality Improvement Project Investigators. Using quality of life to predict hospitalization and mortality in patients with obstructive lung diseases. Chest 2002 Aug;122(2):429-436. [doi: 10.1378/chest.122.2.429] [Medline: 12171813]
44. Moy ML, Teylan M, Danilack VA, Gagnon DR, Garshick E. An index of daily step count and systemic inflammation predicts clinical outcomes in chronic obstructive pulmonary disease. Ann Am Thorac Soc 2014 Feb;11(2):149-157. [doi: 10.1513/AnnalsATS.201307-243OC] [Medline: 24308588]
45. Briggs A, Spencer M, Wang H, Mannino D, Sin DD. Development and validation of a prognostic index for health outcomes in chronic obstructive pulmonary disease. Arch Intern Med 2008 Jan 14;168(1):71-79. [doi: 10.1001/archinternmed.2007.37] [Medline: 18195198]
46. Lange P, Marott JL, Vestbo J, Olsen KR, Ingebrigtsen TS, Dahl M, et al. Prediction of the clinical course of chronic obstructive pulmonary disease, using the new GOLD classification: a study of the general population. Am J Respir Crit Care Med 2012 Nov 15;186(10):975-981. [doi: 10.1164/rccm.201207-1299OC] [Medline: 22997207]
47. Abascal-Bolado B, Novotny PJ, Sloan JA, Karpman C, Dulohery MM, Benzo RP. Forecasting COPD hospitalization in the clinic: optimizing the chronic respiratory questionnaire. Int J Chron Obstruct Pulmon Dis 2015;10:2295-2301 [FREE Full text] [doi: 10.2147/COPD.S87469] [Medline: 26543362]
48. Blanco-Aparicio M, Vázquez I, Pita-Fernández S, Pértega-Diaz S, Verea-Hernando H. Utility of brief questionnaires of health-related quality of life (Airways Questionnaire 20 and Clinical COPD Questionnaire) to predict exacerbations in patients with asthma and COPD. Health Qual Life Outcomes 2013 May 27;11:85 [FREE Full text] [doi: 10.1186/1477-7525-11-85] [Medline: 23706146]
49. Yoo JW, Hong Y, Seo JB, Chae EJ, Ra SW, Lee JH, et al. Comparison of clinico-physiologic and CT imaging risk factors for COPD exacerbation. J Korean Med Sci 2011 Dec;26(12):1606-1612 [FREE Full text] [doi: 10.3346/jkms.2011.26.12.1606] [Medline: 22147998]
50. Niewoehner DE, Lokhnygina Y, Rice K, Kuschner WG, Sharafkhaneh A, Sarosi GA, et al. Risk indexes for exacerbations and hospitalizations due to COPD. Chest 2007 Jan;131(1):20-28. [doi: 10.1378/chest.06-1316] [Medline: 17218552]
51. Austin PC, Stanbrook MB, Anderson GM, Newman A, Gershon AS. Comparative ability of comorbidity classification methods for administrative data to predict outcomes in patients with chronic obstructive pulmonary disease. Ann Epidemiol 2012 Dec;22(12):881-887 [FREE Full text] [doi: 10.1016/j.annepidem.2012.09.011] [Medline: 23121992]
52. Marin JM, Carrizo SJ, Casanova C, Martinez-Camblor P, Soriano JB, Agusti AG, et al. Prediction of risk of COPD exacerbations by the BODE index. Respir Med 2009 Mar;103(3):373-378 [FREE Full text] [doi: 10.1016/j.rmed.2008.10.004] [Medline: 19013781]
53. Ställberg B, Lisspers K, Larsson K, Janson C, Müller M, Łuczko M, et al. Predicting hospitalization due to COPD exacerbations in Swedish primary care patients using machine learning - based on the ARCTIC study. Int J Chron Obstruct Pulmon Dis 2021;16:677-688 [FREE Full text] [doi: 10.2147/COPD.S293099] [Medline: 33758504]
54. Tong Y, Liao ZC, Tarczy-Hornoch P, Luo G. Using a constraint-based method to identify chronic disease patients who are apt to obtain care mostly within a given health care system: retrospective cohort study. JMIR Form Res 2021 Oct 07;5(10):e26314 [FREE Full text] [doi: 10.2196/26314] [Medline: 34617906]
55. NQF #1891 Hospital 30-day, all-cause, risk-standardized readmission rate (RSRR) following chronic obstructive pulmonary disease (COPD) hospitalization. National Quality Forum. 2012. URL: http://www.qualityforum.org/Projects/n-r/Pulmonary_Endorsement_Maintenance/1891_30_Day_RSRR_COPD.aspx [accessed 2021-12-19]
56. Cooke CR, Joo MJ, Anderson SM, Lee TA, Udris EM, Johnson E, et al. The validity of using ICD-9 codes and pharmacy records to identify patients with chronic obstructive pulmonary disease. BMC Health Serv Res 2011 Feb 16;11:37 [FREE Full text] [doi: 10.1186/1472-6963-11-37] [Medline: 21324188]
57. Nguyen HQ, Chu L, Amy Liu IL, Lee JS, Suh D, Korotzer B, et al. Associations between physical activity and 30-day readmission risk in chronic obstructive pulmonary disease. Ann Am Thorac Soc 2014 Jun;11(5):695-705. [doi: 10.1513/AnnalsATS.201401-017OC] [Medline: 24713094]
58. Lindenauer PK, Grosso LM, Wang C, Wang Y, Krishnan JA, Lee TA, et al. Development, validation, and results of a risk-standardized measure of hospital 30-day mortality for patients with exacerbation of chronic obstructive pulmonary disease. J Hosp Med 2013 Aug;8(8):428-435. [doi: 10.1002/jhm.2066] [Medline: 23913593]
59. Qureshi H, Sharafkhaneh A, Hanania NA. Chronic obstructive pulmonary disease exacerbations: latest evidence and clinical implications. Ther Adv Chronic Dis 2014 Sep;5(5):212-227 [FREE Full text] [doi: 10.1177/2040622314532862] [Medline: 25177479]

60. Müllerova H, Maselli DJ, Locantore N, Vestbo J, Hurst JR, Wedzicha JA, et al. Hospitalized exacerbations of COPD: risk factors and outcomes in the ECLIPSE cohort. Chest 2015 Apr;147(4):999-1007. [doi: 10.1378/chest.14-0655] [Medline: 25356881]
61. Donaldson GC, Seemungal TA, Bhowmik A, Wedzicha JA. Relationship between exacerbation frequency and lung function decline in chronic obstructive pulmonary disease. Thorax 2002 Oct;57(10):847-852 [FREE Full text] [doi: 10.1136/thorax.57.10.847] [Medline: 12324669]
62. Hurst JR, Donaldson GC, Quint JK, Goldring JJ, Baghai-Ravary R, Wedzicha JA. Temporal clustering of exacerbations in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2009 Mar 01;179(5):369-374. [doi: 10.1164/rccm.200807-1067OC] [Medline: 19074596]
63. Similowski T, Agustí A, MacNee W, Schönhofer B. The potential impact of anaemia of chronic disease in COPD. Eur Respir J 2006 Feb;27(2):390-396 [FREE Full text] [doi: 10.1183/09031936.06.00143704] [Medline: 16452598]
64. Dahl M, Vestbo J, Lange P, Bojesen SE, Tybjaerg-Hansen A, Nordestgaard BG. C-reactive protein as a predictor of prognosis in chronic obstructive pulmonary disease. Am J Respir Crit Care Med 2007 Feb 1;175(3):250-255. [doi: 10.1164/rccm.200605-713OC] [Medline: 17053205]
65. Hoenderdos K, Condliffe A. The neutrophil in chronic obstructive pulmonary disease. Am J Respir Cell Mol Biol 2013 May;48(5):531-539. [doi: 10.1165/rcmb.2012-0492TR] [Medline: 23328639]
66. Lonergan M, Dicker AJ, Crichton ML, Keir HR, Van Dyke MK, Mullerova H, et al. Blood neutrophil counts are associated with exacerbation frequency and mortality in COPD. Respir Res 2020 Jul 01;21(1):166 [FREE Full text] [doi: 10.1186/s12931-020-01436-7] [Medline: 32611352]
67. Chambellan A, Chailleux E, Similowski T, ANTADIR Observatory Group. Prognostic value of the hematocrit in patients with severe COPD receiving long-term oxygen therapy. Chest 2005 Sep;128(3):1201-1208. [doi: 10.1378/chest.128.3.1201] [Medline: 16162707]
68. Toft-Petersen AP, Torp-Pedersen C, Weinreich UM, Rasmussen BS. Association between hemoglobin and prognosis in patients admitted to hospital for COPD. Int J Chron Obstruct Pulmon Dis 2016;11:2813-2820 [FREE Full text] [doi: 10.2147/COPD.S116269] [Medline: 27877035]
69. van Dijk EJ, Vermeer SE, de Groot JC, van de Minkelis J, Prins ND, Oudkerk M, et al. Arterial oxygen saturation, COPD, and cerebral small vessel disease. J Neurol Neurosurg Psychiatry 2004 May;75(5):733-736 [FREE Full text] [doi: 10.1136/jnnp.2003.022012] [Medline: 15090569]
70. Kessler R, Faller M, Fourgaut G, Mennecier B, Weitzenblum E. Predictive factors of hospitalization for acute exacerbation in a series of 64 patients with chronic obstructive pulmonary disease. Am J Respir Crit Care Med 1999 Jan;159(1):158-164. [doi: 10.1164/ajrccm.159.1.9803117] [Medline: 9872834]
71. Fermont JM, Masconi KL, Jensen MT, Ferrari R, Di Lorenzo VA, Marott JM, et al. Biomarkers and clinical outcomes in COPD: a systematic review and meta-analysis. Thorax 2019 May;74(5):439-446 [FREE Full text] [doi: 10.1136/thoraxjnl-2018-211855] [Medline: 30617161]
72. Halpin DM, Miravitlles M, Metzdorf N, Celli B. Impact and prevention of severe exacerbations of COPD: a review of the evidence. Int J Chron Obstruct Pulmon Dis 2017;12:2891-2908 [FREE Full text] [doi: 10.2147/COPD.S139470] [Medline: 29062228]
73. The world's oldest people and their secrets to a long life. Guinness World Records. 2020. URL: https://www.guinnessworldrecords.com/news/2020/10/the-worlds-oldest-people-and-their-secrets-to-a-long-life-632895 [accessed 2021-12-20]
74. Lightest birth. Guinness World Records. 2020. URL: https://www.guinnessworldrecords.com/world-records/lightest-birth [accessed 2021-12-20]
75. Heaviest man ever. Guinness World Records. 2020. URL: https://www.guinnessworldrecords.com/world-records/heaviest-man [accessed 2021-12-20]
76. Shortest baby. Guinness World Records. 2020. URL: https://www.guinnessworldrecords.com/world-records/shortest-baby [accessed 2021-12-20]
77. Tallest man ever. Guinness World Records. 2020. URL: https://www.guinnessworldrecords.com/world-records/tallest-man-ever [accessed 2021-12-20]
78. Gwyneth O. Part V Fat: no more fear, no more contempt. The Eating Disorder Institute. 2011. URL: https://edinstitute.org/blog/2011/12/8/part-v-fat-no-more-fear-no-more-contempt [accessed 2021-12-20]
79. List of heaviest people. Wikipedia. 2021. URL: https://en.wikipedia.org/w/index.php?title=List_of_heaviest_people&oldid=1000662342 [accessed 2021-12-20]
80. Hankinson JL, Odencrantz JR, Fedan KB. Spirometric reference values from a sample of the general U.S. population. Am J Respir Crit Care Med 1999 Jan;159(1):179-187. [doi: 10.1164/ajrccm.159.1.9712108] [Medline: 9872837]
81. Pellegrino R, Viegi G, Brusasco V, Crapo RO, Burgos F, Casaburi R, et al. Interpretative strategies for lung function tests. Eur Respir J 2005 Nov;26(5):948-968 [FREE Full text] [doi: 10.1183/09031936.05.00035205] [Medline: 16264058]
82. Marion MS, Leonardson GR, Rhoades ER, Welty TK, Enright PL. Spirometry reference values for American Indian adults: results from the Strong Heart Study. Chest 2001 Aug;120(2):489-495. [doi: 10.1378/chest.120.2.489] [Medline: 11502648]

83. Bronchodilators. National Jewish Health. 2018. URL: https://nationaljewish.org/conditions/medications/copd/bronchodilators [accessed 2021-12-19]
84. Luo G, He S, Stone BL, Nkoy FL, Johnson MD. Developing a model to predict hospital encounters for asthma in asthmatic patients: secondary analysis. JMIR Med Inform 2020 Jan 21;8(1):e16080 [FREE Full text] [doi: 10.2196/16080] [Medline: 31961332]
85. Tong Y, Messinger AI, Wilcox AB, Mooney SD, Davidson GH, Suri P, et al. Forecasting future asthma hospital encounters of patients with asthma in an academic health care system: predictive model development and secondary analysis study. J Med Internet Res 2021 Apr 16;23(4):e22796 [FREE Full text] [doi: 10.2196/22796] [Medline: 33861206]
86. Witten IH, Frank E, Hall MA, Pal CJ. Data Mining: Practical Machine Learning Tools and Techniques, 4th ed. Burlington, MA: Morgan Kaufmann; 2016.
87. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016 Presented at: KDD'16; August 13-17, 2016; San Francisco, CA p. 785-794. [doi: 10.1145/2939672.2939785]
88. XGBoost JVM package. 2021. URL: https://xgboost.readthedocs.io/en/latest/jvm/index.html [accessed 2021-12-20]
89. Zeng X, Luo G. Progressive sampling-based Bayesian optimization for efficient and automatic machine learning model selection. Health Inf Sci Syst 2017 Dec;5(1):2 [FREE Full text] [doi: 10.1007/s13755-017-0023-z] [Medline: 29038732]
90. Thornton C, Hutter F, Hoos HH, Leyton-Brown K. Auto-WEKA: combined selection and hyperparameter optimization of classification algorithms. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2013 Presented at: KDD'13; August 11-14, 2013; Chicago, IL p. 847-855. [doi: 10.1145/2487575.2487629]
91. Steyerberg EW. Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating, 2nd ed. New York, NY: Springer; 2019.
92. Sykes DL, Faruqi S, Holdsworth L, Crooks MG. Impact of COVID-19 on COPD and asthma admissions, and the pandemic from a patient's perspective. ERJ Open Res 2021 Feb 8;7(1) [FREE Full text] [doi: 10.1183/23120541.00822-2020] [Medline: 33575313]
93. Agresti A. Categorical Data Analysis, 3rd ed. Hoboken, NJ: Wiley; 2012.
94. Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed. New York, NY: Springer; 2016.
95. Luo G, Johnson MD, Nkoy FL, He S, Stone BL. Automatically explaining machine learning prediction results on asthma hospital visits in patients with asthma: secondary analysis. JMIR Med Inform 2020 Dec 31;8(12):e21965 [FREE Full text] [doi: 10.2196/21965] [Medline: 33382379]
96. Guerra B, Gaveikaite V, Bianchi C, Puhan MA. Prediction models for exacerbations in patients with COPD. Eur Respir Rev 2017 Jan 17;26(143):160061 [FREE Full text] [doi: 10.1183/16000617.0061-2016] [Medline: 28096287]
97. Bellou V, Belbasis L, Konstantinidis AK, Tzoulaki I, Evangelou E. Prognostic models for outcome prediction in patients with chronic obstructive pulmonary disease: systematic review and critical appraisal. Br Med J 2019 Oct 4;367:l5358 [FREE Full text] [doi: 10.1136/bmj.l5358] [Medline: 31585960]
98. Longman JM, Passey ME, Ewald DP, Rix E, Morgan GG. Admissions for chronic ambulatory care sensitive conditions - a useful measure of potentially preventable admission? BMC Health Serv Res 2015 Oct 16;15:472 [FREE Full text] [doi: 10.1186/s12913-015-1137-0] [Medline: 26475293]
99. Johnston JJ, Longman JM, Ewald DP, Rolfe MI, Alvarez SD, Gilliland AH, et al. Validity of a tool designed to assess the preventability of potentially preventable hospitalizations for chronic conditions. Fam Pract 2020 Jul 23;37(3):390-394 [FREE Full text] [doi: 10.1093/fampra/cmz086] [Medline: 31848589]
100. Ranganathan P, Aggarwal R. Common pitfalls in statistical analysis: understanding the properties of diagnostic tests - Part 1. Perspect Clin Res 2018;9(1):40-43 [FREE Full text] [doi: 10.4103/picr.PICR_170_17] [Medline: 29430417]
101. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. In: Proceedings of the 23rd International Conference on Machine Learning. 2006 Presented at: ICML'06; June 25-29, 2006; Pittsburgh, PA p. 233-240. [doi: 10.1145/1143844.1143874]
102. Burge AT, Holland AE, McDonald CF, Abramson MJ, Hill CJ, Lee AL, et al. Home-based pulmonary rehabilitation for COPD using minimal resources: an economic analysis. Respirology 2020 Feb;25(2):183-190. [doi: 10.1111/resp.13667] [Medline: 31418515]
103. XGBoost Parameters. 2021. URL: https://xgboost.readthedocs.io/en/latest/parameter.html [accessed 2021-12-20]
104. Notes on Parameter Tuning. 2021. URL: https://xgboost.readthedocs.io/en/latest/tutorials/param_tuning.html [accessed 2021-12-20]
105. Stein BD, Bautista A, Schumock GT, Lee TA, Charbeneau JT, Lauderdale DS, et al. The validity of International Classification of Diseases, Ninth Revision, Clinical Modification diagnosis codes for identifying patients hospitalized for COPD exacerbations. Chest 2012 Jan;141(1):87-93 [FREE Full text] [doi: 10.1378/chest.11-0024] [Medline: 21757568]
106. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018 May 8;1:18 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]

107. Lipton ZC, Kale DC, Elkan C, Wetzel RC. Learning to diagnose with LSTM recurrent neural networks. In: Proceedings of the International Conference on Learning Representations. 2016 Presented at: International Conference on Learning Representations; May 2-4, 2016; San Juan, Puerto Rico p. 1-18 URL: https://arxiv.org/abs/1511.03677
108. Kam HJ, Kim HY. Learning representations for the early detection of sepsis with deep neural networks. Comput Biol Med 2017 Dec 01;89:248-255. [doi: 10.1016/j.compbiomed.2017.08.015] [Medline: 28843829]
109. Razavian N, Marcus J, Sontag D. Multi-task prediction of disease onsets from longitudinal laboratory tests. In: Proceedings of the Machine Learning in Health Care Conference. 2016 Presented at: Machine Learning in Health Care Conference; August 19-20, 2016; Los Angeles, CA p. 73-100 URL: http://proceedings.mlr.press/v56/Razavian16.html
110. Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface 2018 Apr;15(141):20170387 [FREE Full text] [doi: 10.1098/rsif.2017.0387] [Medline: 29618526]
111. Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for Electronic Health Record (EHR) analysis. IEEE J Biomed Health Inform 2018 Dec;22(5):1589-1604. [doi: 10.1109/JBHI.2017.2767063] [Medline: 29989977]
112. Luo G, Stone BL, Koebnick C, He S, Au DH, Sheng X, et al. Using temporal features to provide data-driven clinical early warnings for chronic obstructive pulmonary disease and asthma care management: protocol for a secondary analysis. JMIR Res Protoc 2019 Jun 06;8(6):e13783 [FREE Full text] [doi: 10.2196/13783] [Medline: 31199308]

Abbreviations

AUC: area under the receiver operating characteristic curve
AUPRC: area under the precision–recall curve
COPD: chronic obstructive pulmonary disease
ED: emergency department
LABA: long-acting beta-2 agonist
LAMA: long-acting muscarinic antagonist
NPV: negative predictive value
PCP: primary care provider
PPV: positive predictive value
UWM: University of Washington Medicine
WEKA: Waikato Environment for Knowledge Analysis
XGBoost: Extreme Gradient Boosting

Edited by G Eysenbach; submitted 03.04.21; peer-reviewed by V Press, P Orchard; comments to author 28.06.21; revised version received 03.07.21; accepted 19.11.21; published 06.01.22.

Please cite as:
Zeng S, Arjomandi M, Tong Y, Liao ZC, Luo G
Developing a Machine Learning Model to Predict Severe Chronic Obstructive Pulmonary Disease Exacerbations: Retrospective Cohort Study
J Med Internet Res 2022;24(1):e28953
URL: https://www.jmir.org/2022/1/e28953
doi:10.2196/28953
PMID:34989686

Original Paper

Implementation and Evaluation of a Digitally Enabled Precision Public Health Intervention to Reduce Inappropriate Gabapentinoid Prescription: Cluster Randomized Controlled Trial

Andre Q Andrade[1], MD, PhD; Jean-Pierre Calabretto[1,2], PhD; Nicole L Pratt[1], PhD; Lisa M Kalisch-Ellett[1], PhD; Gizat M Kassie[1], PhD; Vanessa T LeBlanc[1]; Emmae Ramsay[1]; Elizabeth E Roughead[1], PhD

[1]Quality Use of Medicines and Pharmacy Research Centre, UniSA Clinical and Medical Sciences, University of South Australia, Adelaide, Australia
[2]Australian Medicines Handbook Pty Ltd, Adelaide, Australia

Corresponding Author:
Andre Q Andrade, MD, PhD
Quality Use of Medicines and Pharmacy Research Centre
UniSA Clinical and Medical Sciences
University of South Australia
GPO Box 2471
Adelaide, 5001
Australia
Phone: 61 8 8302 2314
Email: andre.andrade@unisa.edu.au

Abstract

Background: Digital technologies can enable rapid targeted delivery of audit and feedback interventions at scale. Few studies have evaluated how mode of delivery affects clinical professional behavior change and none have assessed the feasibility of such an initiative at a national scale.

Objective: The aim of this study was to develop and evaluate the effect of audit and feedback by digital versus postal (letter) mode of delivery on primary care physician behavior.

Methods: This study was developed as part of the Veterans' Medicines Advice and Therapeutics Education Services (MATES) program, an intervention funded by the Australian Government Department of Veterans' Affairs that provides targeted education and patient-specific audit with feedback to Australian general practitioners, as well as educational material to veterans and other health professionals. We performed a cluster randomized controlled trial of a multifaceted intervention to reduce inappropriate gabapentinoid prescription, comparing digital and postal mode of delivery. All veteran patients targeted also received an educational intervention (postal delivery). Efficacy was measured using a linear mixed-effects model as the average number of gabapentinoid prescriptions standardized by defined daily dose (individual level), and number of veterans visiting a psychologist in the 6 and 12 months following the intervention.

Results: The trial involved 2552 general practitioners in Australia and took place in March 2020. Both intervention groups had a significant reduction in total gabapentinoid prescription by the end of the study period (digital: mean reduction of 11.2%, *P*=.004; postal: mean reduction of 11.2%, *P*=.001). We found no difference between digital and postal mode of delivery in reduction of gabapentinoid prescriptions at 12 months (digital: –0.058, postal: –0.058, *P*=.98). Digital delivery increased initiations to psychologists at 12 months (digital: 3.8%, postal: 2.0%, *P*=.02).

Conclusions: Our digitally delivered professional behavior change intervention was feasible, had comparable effectiveness to the postal intervention with regard to changes in medicine use, and had increased effectiveness with regard to referrals to a psychologist. Given the logistical benefits of digital delivery in nationwide programs, the results encourage exploration of this mode in future interventions.

(J Med Internet Res 2022;24(1):e33873) doi:10.2196/33873

KEYWORDS

audit and feedback; digital health; precision public health; digital intervention; primary care; physician; health professional; health education

Introduction

Audit and feedback interventions can be effective tools to promote evidence translation through professional behavior change [1]. Audit and feedback interventions objectively measure professional performance and create benchmarks against professional standards. Development and dissemination of audit and feedback interventions have benefited from advances in information technology that have increased data availability, decreased costs, and improved automation. As a cost-effective and data-driven intervention, audit and feedback seems well suited for migrating to a fully electronic mode of delivery [2].

Despite the potential advantages of electronic delivery, there is a theoretical and evidence gap regarding the influence of changing the mode of delivery on the efficacy of behavior change interventions [3]. Previous studies of behavior change interventions suggest that the mode of delivery may influence the efficacy of behavior change techniques. The most likely mechanism is a fundamental change in user experience, which may elicit different responses [3]. Use of different modes of delivery changes users' experiences by creating new contexts (eg, SMS text message sent at any given time versus scheduled educational sessions), creating more personal experiences (eg, face-to-face group sessions versus social media), and providing new modes of interaction (eg, interactive computer interventions versus printed material). A review on the use of behavior change techniques for smoking cessation found a positive effect of the techniques when delivered in person, but not when delivered in writing [4]. Another review on internet-delivered behavior change interventions found increased efficacy when using additional modes of delivery, such as SMS text messages and email communication [5]. However, further analysis of the same data set did not find a synergistic effect between any combination of mode of delivery and behavior change technique. The authors suggest that having additional channels of delivery may be beneficial, but were unable to recommend which modes to use for particular behavior change techniques [6].

Evidence on the influence of digital delivery in audit and feedback's efficacy is also needed. A 2017 review of electronic audit and feedback interventions found heterogeneous results due to differences in the intervention implementation and the underlying theory and context [7].

Following the suggestions put forward by [8], the aim of this study was to evaluate the influence of delivering an audit and feedback intervention by secure digital delivery to the clinical desktop for integration to the patient care record, and compare it to the same intervention delivered by post. The behavior change goal was the reduction in gabapentinoid prescription. Gabapentinoids are a group of medicines that includes gabapentin and pregabalin. Evidence suggests these medicines are often incorrectly prescribed in nonneuropathic pain [9], with significant risk of serious side effects and potential for abuse and misuse [10]. To test the efficacy of the digital intervention, we (the authors) performed a cluster randomized trial of an intervention aimed at reducing inappropriate prescription of gabapentinoids by primary care providers.

Methods

The Veterans' Medicines Advice and Therapeutics Education Services Program

The study is part of the Veterans' Medicines Advice and Therapeutics Education Services (MATES) program [11], which is funded by the Australian Government Department of Veterans' Affairs and provides medicines advice and promotes physician adoption of best practices. Since 2004, it has provided repeated multifaceted interventions, composed of an audit and feedback and educational component targeted at general practitioners (GPs), with supportive educational material provided to veterans, pharmacists, and other health professionals. The intervention is informed by social cognitive theory [12], the transtheoretical model [13], and the health promotion model PRECEDE-PROCEED [14]. Between 2004 and 2021, the program delivered 62 distinct interventions to GPs and veterans in all Australian states.

The intervention is developed in three sequential steps. The first is an epidemiological analysis performed on a comprehensive database containing administrative health claims data collected by the Australian Government Department of Veterans' Affairs (DVA). The DVA claims database includes all health care services and medicines funded by DVA, including outpatient and hospital services, aged care, prescription medicines, allied health services, and other health coordination and support services.

The second step is the design of the educational component. It involves clinicians, researchers, and veterans, and results in two sets of educational materials. The first is targeted at GPs, and describes scientific updates and therapeutic recommendations. The second is targeted at veterans, and promotes general awareness and practical guidelines for patients.

The third step is the development of the audit and feedback component. The design process is also collaborative. The intervention adopts evidence-based strategies listed in [15] to improve effectiveness, such as authority (content endorsed by a clinical DVA committee), focus on problems with larger scope for improvement, and repeated feedback (topics are revisited after a few years). It also incorporates behavior change techniques such as heuristic techniques, goal setting, and prompts, which have been shown to improve perceived usefulness [16].

Digital Solution Design

The digital solution was conducted using a collaborative, pragmatic approach, influenced by Greenhalgh et al's [17] Diffusion of Innovations Model, to develop a solution that could be implemented at a national scale. We used a series of stakeholder meetings scheduled as part of the Veterans' MATES program to map out context and understand adopters' (ie, GPs') practices and preferences. The meetings involved funder (DVA) representatives, clinicians, veteran representatives, and information technology professionals.

To develop the solution, some particularities of the Australian health system and context were considered relevant, in particular:

- Reliance on primary care providers: The GP is the gatekeeper of the Australian health care system. About 84% of Australians see a GP every year, and 77% of patients have a preferred GP [18].
- Geographical distribution: GPs responsible for Australian veterans are located in all parts of the country. There are only a few GPs specialized in veteran care; most professionals have less than a handful of veterans under their care.
- Technological readiness in primary care: Use of electronic health records in primary care has been widespread in Australia for at least 10 years [19]. Additionally, secure messaging infrastructure is well established for receiving laboratory test results.

The proposed solution is an adaptation of the 3 steps used by Veterans' MATES interventions, suited for a digital medium. To identify individuals at risk of medication-related harm, the solution uses a set of algorithms to extract information from claims data (services and medicines) indicating phenotype (which conditions affect the patient based on the resources they use). These algorithms identify patients at risk, either due to long-term conditions, medicine use, or current events such as medicine discontinuation.

To create the electronic messages, patient information extracted from the claims database is embedded in a template to create an audit and feedback document designed to promote recognition of patient risk. The document uses behavior change techniques, including prompts, goal setting, discrepancy between current behavior and goal, information about health consequences, and feedback on behavior; all of these techniques have been shown to improve intervention usefulness [16].

Documents are created as PDF documents, encrypted, and embedded in a Health Level Seven (HL7) version 2 file using internally developed software. Audit and feedback documents may contain complex graphical elements and may change significantly according to patients' conditions and suggested recommendations. Therefore, we chose to initially develop documents as HTML pages, which are then converted to PDF format.

Finally, our investigation suggested the suitability of using an existing secure message infrastructure to reach GPs. Encrypted HL7 messages are sent to GP offices using a third-party provider and then decrypted by the clinical software and incorporated into the GP workflow. This solution adheres to many determinants of innovation diffusion identified in [17], as shown in Table 1.

Table 1. Determinants of innovation diffusion and predicted advantages of proposed solution.

Determinant of innovation	Predicted advantages of proposed solution
Relative advantage	Electronic messages are easier to read and act upon and less cumbersome than other communication means, such as printed materials or telephone communication
Compatibility	The solution uses communication infrastructure already being used to receive laboratory test results, with minimal additional impact on clinician workflow
Complexity	The solution can be described by the three main processes (patient identification, message tailoring, and secure delivery), which are understood by all stakeholders
Trialability	The solution was trialed in 3 small pilots and 1 randomized controlled trial before large-scale adoption
Risk	The solution has a relatively low cost and builds upon a 15-year program, reducing risk
Task issues	The solution is embedded in current workflow, with minimal task disruption
Augmentation/support	Each message is data driven, meaning it offers information related to a unique patient, also providing clear and unambiguous recommendations

Feasibility Studies

The most important implementation risk identified during the initial stakeholder meetings was that the intervention could be perceived as intrusive and disruptive to GP workflow. To mitigate this risk, the solution was trialed in 3 sequential small-scale pilots, taking place in April, July, and September 2019. The main goals of the pilots were the following:

1. Evaluate the technical feasibility. We measured the proportion of messages acknowledged as successfully received.
2. Reduce the risk of disrupting GP work practices. GPs involved in the pilot could get in touch via a support email, website, and telephone. Additionally, we sent an invitation to an online survey containing 16 questions about usability and satisfaction.

The first pilot was planned as an opt-in trial, and GPs were invited to participate by email (Multimedia Appendix 1). The second pilot was planned as an opt-out trial, and GPs were sent an email explaining the study and offering the opportunity to be removed from the list. The final trial was planned as usual service and preceded by a mailed information leaflet (Multimedia Appendix 2).

Trial Design

To test the influence of mode of delivery on the effectiveness of audit and feedback interventions, we performed a parallel, cluster randomized trial of a computer-delivered intervention to reduce inappropriate gabapentinoid prescription. The trial

was designed to compare the post-delivered intervention (usual intervention as concurrent control) with the computer-delivered intervention. Since the intervention targets GPs who may have multiple veteran patients, we adopted a cluster design whereby a GP received information for all of their patients in the intervention by the same mode.

The intervention delivered via postal mode has been shown effective in translating evidence in different domains [20] including promoting medicine review [21], osteoporosis screening [22], uptake of health services [23], reducing inappropriate proton pump inhibitor use [24], and hypnotic use for insomnia [25].

Participants

To be eligible for participation, both veterans and their primary GP had to be eligible for the digital intervention. Eligible veterans comprised active DVA clients that had 2 or more gabapentinoid (either pregabalin or gabapentin) prescriptions in a 4-month period (October 2019 to January 2020). Veterans were also required to be resident in Australia, living in the community setting (ie, not residing in aged care or other long-term care facilities), and to not have previously requested exclusion from Veterans' MATES interventions for any reason.

GPs were eligible if they were identified as the primary GP of one or more Australian veterans, and at least one of the veterans was eligible for the intervention. Participant GPs were excluded if they did not have installed capacity to receive secure electronic messages from our partner message provider (HealthLink Group Limited) or if they had previously requested exclusion from Veterans' MATES interventions for any reason.

To determine the primary GP for a given veteran, we developed an algorithm based on prescriptions and outpatient services provided. Providers were scored based on the number of prescriptions and services provided, and weighted based on recency of services to account for veterans changing providers.

Setting

The trial was conducted across all Australian states and territories. We determined patient eligibility by querying the DVA claims database. Outcome data including service provision and medicine dispensing were collected from the DVA claims database.

The Intervention

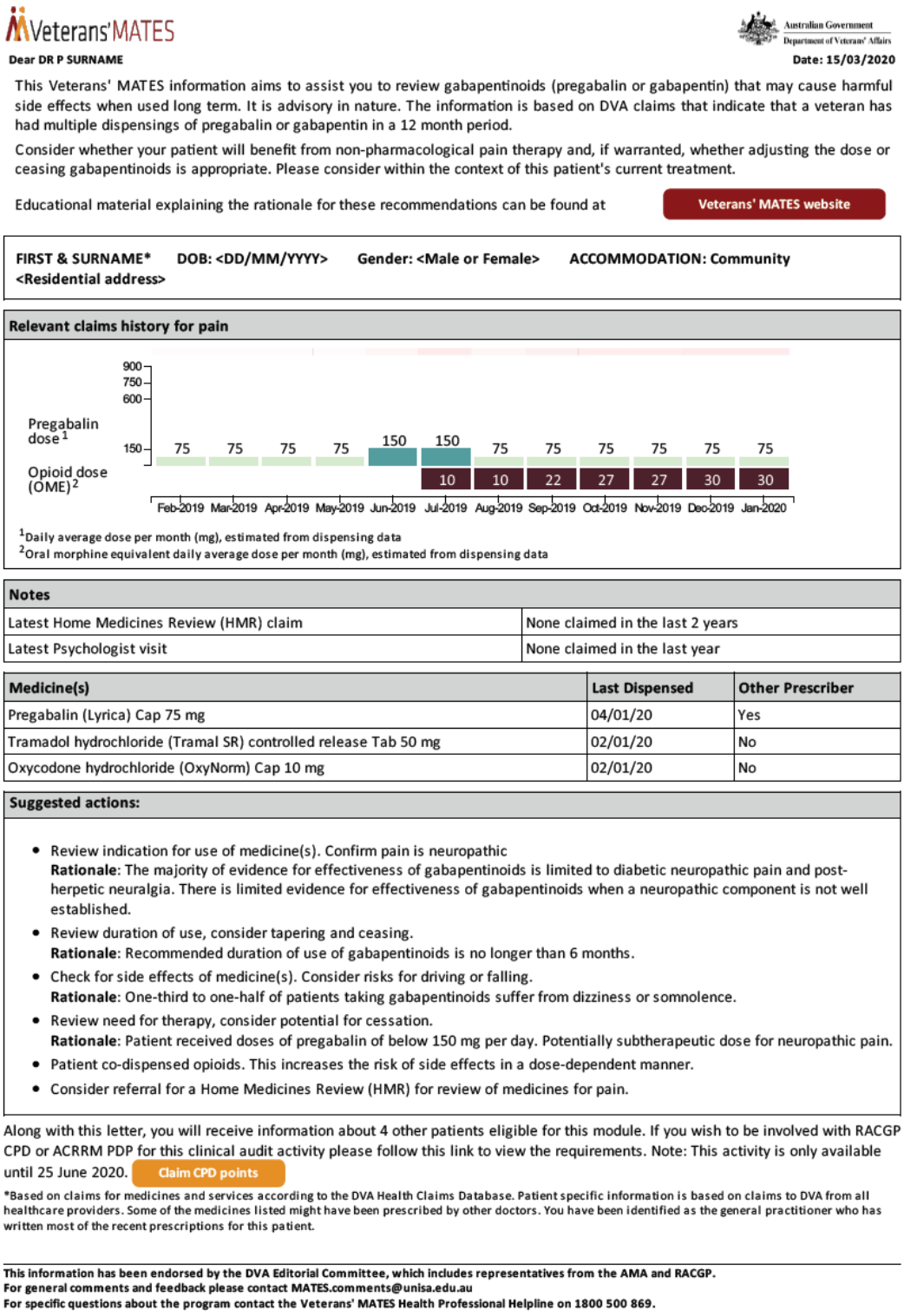
GPs in the intervention arm received the intervention exclusively in a digitally delivered format. It was sent via secure message infrastructure directly to the GP's clinic. Once received by the practice, it is reviewed by a practice manager of the GP and assigned to the appropriate patient. Once it is assigned, it can be accessed in the electronic health record alongside pathology reports and referral letters.

GPs in the usual care arm received the intervention by postal service. This delivery contains both the audit and feedback documents (for all selected veteran patients) and the educational materials (including a copy of the material targeted at veterans).

Both sets of materials contained the same theoretical content and personal information. Due to feedback from users, the digitally delivered intervention was slightly modified to user workflow. Since we could not deliver general educational documents to the health record, the audit and feedback document was enhanced to contain a link to the online educational material (see Figure 1). Additionally, the single letter containing multiple patients was segmented into one electronic document per patient. Finally, a color chart was added at the top of the electronic document to highlight different prescription patterns and help GPs prioritize patients when receiving multiple documents.

Veterans in both the intervention and usual care arms received educational material by post. The material can be found on the Veterans' MATES web page [26].

Figure 1. Example of the intervention delivered to general practitioners (digital version).



Veterans' MATES

Australian Government Department of Veterans' Affairs

Dear DR P SURNAME **Date: 15/03/2020**

This Veterans' MATES information aims to assist you to review gabapentinoids (pregabalin or gabapentin) that may cause harmful side effects when used long term. It is advisory in nature. The information is based on DVA claims that indicate that a veteran has had multiple dispensings of pregabalin or gabapentin in a 12 month period.

Consider whether your patient will benefit from non-pharmacological pain therapy and, if warranted, whether adjusting the dose or ceasing gabapentinoids is appropriate. Please consider within the context of this patient's current treatment.

Educational material explaining the rationale for these recommendations can be found at Veterans' MATES website

FIRST & SURNAME* **DOB: <DD/MM/YYYY>** **Gender: <Male or Female>** **ACCOMMODATION: Community**
<Residential address>

Relevant claims history for pain

[1]Daily average dose per month (mg), estimated from dispensing data
[2]Oral morphine equivalent daily average dose per month (mg), estimated from dispensing data

Notes	
Latest Home Medicines Review (HMR) claim	None claimed in the last 2 years
Latest Psychologist visit	None claimed in the last year

Medicine(s)	Last Dispensed	Other Prescriber
Pregabalin (Lyrica) Cap 75 mg	04/01/20	Yes
Tramadol hydrochloride (Tramal SR) controlled release Tab 50 mg	02/01/20	No
Oxycodone hydrochloride (OxyNorm) Cap 10 mg	02/01/20	No

Suggested actions:

- Review indication for use of medicine(s). Confirm pain is neuropathic
  Rationale: The majority of evidence for effectiveness of gabapentinoids is limited to diabetic neuropathic pain and post-herpetic neuralgia. There is limited evidence for effectiveness of gabapentinoids when a neuropathic component is not well established.
- Review duration of use, consider tapering and ceasing.
  Rationale: Recommended duration of use of gabapentinoids is no longer than 6 months.
- Check for side effects of medicine(s). Consider risks for driving or falling.
  Rationale: One-third to one-half of patients taking gabapentinoids suffer from dizziness or somnolence.
- Review need for therapy, consider potential for cessation.
  Rationale: Patient received doses of pregabalin of below 150 mg per day. Potentially subtherapeutic dose for neuropathic pain.
- Patient co-dispensed opioids. This increases the risk of side effects in a dose-dependent manner.
- Consider referral for a Home Medicines Review (HMR) for review of medicines for pain.

Along with this letter, you will receive information about 4 other patients eligible for this module. If you wish to be involved with RACGP CPD or ACRRM PDP for this clinical audit activity please follow this link to view the requirements. Note: This activity is only available until 25 June 2020. Claim CPD points

*Based on claims for medicines and services according to the DVA Health Claims Database. Patient specific information is based on claims to DVA from all healthcare providers. Some of the medicines listed might have been prescribed by other doctors. You have been identified as the general practitioner who has written most of the recent prescriptions for this patient.

This information has been endorsed by the DVA Editorial Committee, which includes representatives from the AMA and RACGP.
For general comments and feedback please contact MATES.comments@unisa.edu.au
For specific questions about the program contact the Veterans' MATES Health Professional Helpline on 1800 500 869.

Enrolment and Randomization

Following the eligibility criteria, all GPs acting as the main care provider for an eligible patient (current gabapentinoid use) were considered for recruitment. We excluded all practitioners not found in the partner's (Healthlink Group Limited) provider directory, as they would be unable to receive secure electronic messages. All eligible GP and patient pairs were included in the study sample.

GPs were randomized 1:1 to intervention or usual care. Randomization was block stratified by number of veterans under care. Randomization numbers for each GP were computer generated by a statistician who was not involved in enrolment. Due to the highly automated nature of the intervention and data collection (claims data), no further masking procedures were performed.

An ethics protocol for the study was approved by the University of South Australia Human Research Ethics Committee (ethics

protocol P203/04) and the Australian Government Department of Defence and Veterans' Affairs Human Research Ethics Committee (E016/007).

Outcomes

The primary outcomes were the change in average gabapentinoid prescription during the study period, standardized as multiples of the defined daily dose (DDD) per day, and the proportion of veterans visiting a psychologist for the first time. Primary outcomes were evaluated at 6 and 12 months.

Since the dosing of pregabalin and gabapentin are different, we calculated DDD for each medicine and summed the results. To remove the influence of extreme stockpiling and dispensing data errors, patients with a DDD over 10 (10 times the defined daily dose) were removed from analysis. The DDD was created by the World Health Organization (WHO) as a comparative unit of medicine use [27]. In this study, it allows the comparison of different gabapentinoids, such as gabapentin and pregabalin. The average daily DDD was calculated as per the following formula:

The total mass amount was determined according to all claimed prescriptions of gabapentin and pregabalin in the 3 months prior to the intervention (January 3, 2020, to April 2, 2020) and in the 6 months (July 3, 2020, to October 2, 2020) and 12 months (January 3, 2021, to April 2, 2021) following the intervention.

Secondary outcome was the time to the first visit (face-to-face, telephone, or video) with the primary provider. All outcome measures pertain to the individual (patient).

We conducted a secondary analysis that can be divided into two parts. First, we evaluated the overall intervention impact by measuring changes in average gabapentinoid DDD before and after the intervention. Furthermore, we evaluated whether the dose of gabapentinoid (high, medium, or low) or concurrent use of opioids influenced the efficacy of the different modes of delivery. Veterans were considered to be on a high dose if the average DDD in at least one month of the selection period was >2. Veterans were considered to be on a low dose if the average DDD in every month of the selection period was <0.25. Values between those two values were considered to be a medium dose.

Statistical Methods

We analyzed data from services and medicines claims for all enrolled patients who were alive at 12 months postintervention. To account for the cluster design, we analyzed the primary outcome using a linear mixed-effects model [28], with the GP as the grouping variable. The effect of mode of delivery on patients' likelihood of visiting a psychologist was tested by logistic regression, also using GP as the grouping variable. The time to first GP visit was analyzed by survival analysis. Patients were considered to have the "event" if they had an appointment with the targeted GP. Events were right censored at 3 months (92 days). The relative effect of digital mode versus postal mode was evaluated by a Cox proportional hazards model, with the GP as cluster variable. Secondary analysis was performed by univariate linear mixed-effects model, with the GP as grouping variable. For all hypothesis tests, we considered a 95% CI ($P \leq .05$). All analysis was performed in Python 3.7 (The Python Software Foundation). The main statistical libraries used were Statsmodels (version 0.12) [29] and Lifelines (version 0.25.11) [30].

Availability of Data and Material

The data that support the findings of this study are available from the Australian Government DVA but restrictions apply to the availability of these data, which were used under license for this study, and so are not publicly available.

Results

Feasibility Studies

For the first pilot, a convenience sample of 75 GPs were sent an email invitation to participate (opt-in), and 5 GPs agreed to be included. For the second pilot, we selected a convenience sample of 20 GPs who could opt out of the pilot. For the third pilot, 189 messages were sent to GPs who had not participated previously. We received 6 survey responses, and all responders evaluated the usability as good (easy to read, correct information) and were either likely or very likely to continue to subscribe to future interventions. We received a single letter advising a patient had recently switched medical providers. Given the lack of negative feedback and positive survey responses, the project leadership considered the pilot successful and the intervention feasible, and approved the randomized controlled trial.

Cluster Randomized Controlled Trial

A total of 3271 veterans were considered eligible for the intervention, and 2552 GPs were identified as their main care providers (Figure 2). After randomization, the intervention was successfully delivered in March/April 2020. The postal intervention was sent to GPs on March 19, 2020. The computer-delivered intervention was delivered in three waves (on March 23, March 25, and April 2, 2020).

Veterans randomized to either intervention arm had a similar demographic profile (Table 2). The patterns of gabapentinoid and opioid use were also similar.

By the end of the study, both intervention groups had a significant reduction in gabapentinoid dispensing, as measured by the change in average daily DDD from baseline to 12 months (digital: mean reduction of 0.058, SD 0.38, or 11.2%, P=.004; postal: mean reduction of 0.058, SD 0.37, or 11.2%, P=.001). Figure 3 shows the trends in DDD before and after the intervention.

Figure 2. CONSORT (Consolidated Standards of Reporting Trials) flowchart. GP: general practitioner.

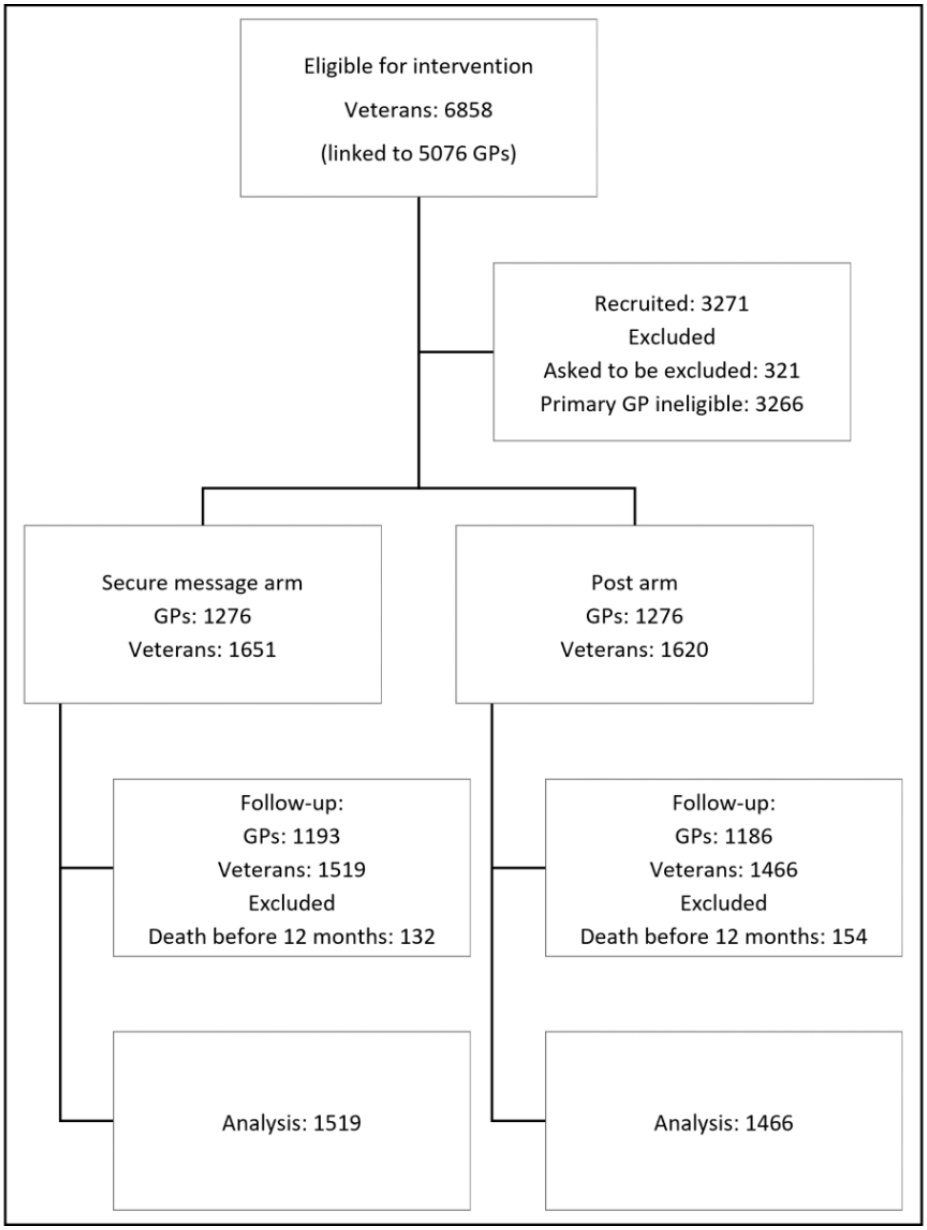
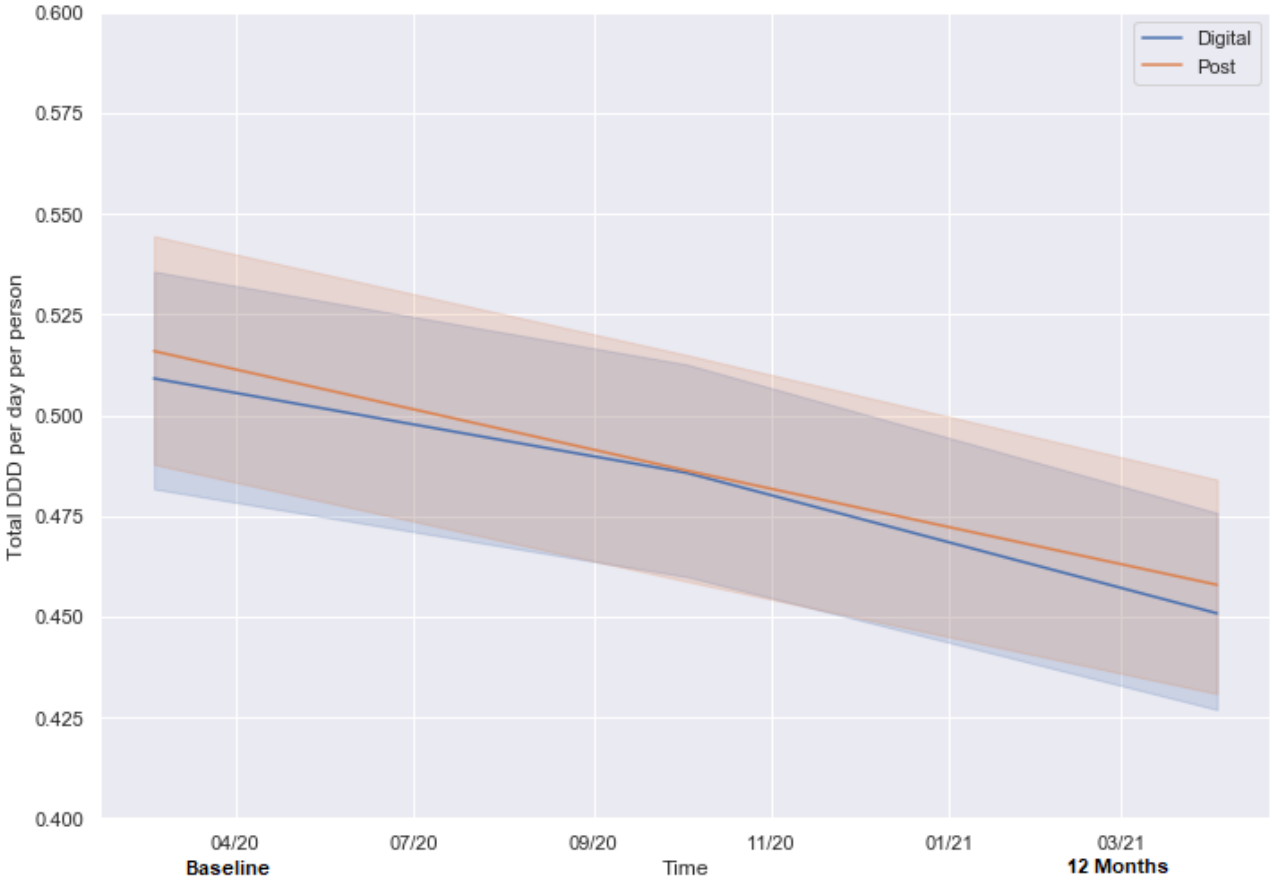


Table 2. Clinical and demographic data at baseline.

Baseline data	Postal intervention	Digital intervention
Number of participants	1466	1519
Age (years), mean (SD)	76.1 (14.6)	76.1 (14.5)
Male, n (%)	853 (58)	883 (58)
Gabapentinoid dose at baseline, n (%)		
High	41 (3)	34 (2)
Medium	1188 (81)	1213 (80)
Low	237 (16)	272 (18)
Concurrent opioid use, n (%)	590 (40)	636 (42)

Figure 3. Average daily DDD by intervention group. DDD: defined daily dose.



We found no difference between digital and postal mode of delivery in reduction of gabapentinoid volume at 6 or 12 months (Table 3). A greater proportion of veterans in the digital intervention group saw a psychologist in the following 12 months (*P*=.02). Digital intervention promoted a small but statistically significant (*P*=.04) effect of earlier GP visits postintervention.

Veterans were segmented according to dose and concurrent opioid use. Consistent with the results of the primary analysis, no differences were found between the digital and postal interventions in any subgroup analysis. Dose reduction was more pronounced in high-dose gabapentinoid users, and there was no observed reduction in the average dose of low-dose users in either arm.

Table 3. Primary and secondary outcomes, by intervention arm.

Outcomes	Postal	Digital	*P* value
Average defined daily dose change (baseline to 6 months)	–0.030	–0.023	.61
Average defined daily dose change (baseline to 12 months)	–0.058	-0.058	.98
Percentage of new psychologist visits (baseline to 6 months)	1.0	1.3	.75
Percentage of new psychologist visits (baseline to 12 months)	2.0	3.8	.02[a]
Hazard ratio for general practitioner visit within 90 days (95% CI)	0.92 (0.85-0.99)	1 (reference)	.04[a]

[a]*P*<.05.

Discussion

Principal Findings

In this study, we present the successful migration of a paper-based national behavior change intervention into a digital intervention. After a careful scaling of the intervention, ample communication, and stakeholder support, we were able to perform a large-scale randomized controlled trial covering all Australian states. The trial showed that both paper and digital versions of an intervention composed of education and audit and feedback was effective in reducing gabapentinoid prescriptions for an Australian population. Additionally, it showed that the digital intervention is equivalent to paper in changing prescription patterns.

This study is one of the first to test the effect of mode of delivery in a large-scale, precision public health intervention. The use of a digital medium of delivery has several advantages over conventional interventions, including the capacity for improved

personalization and precision; improved automation; use of predictive analytics for targeting; data analytics; and improved interaction [31]. However, any new intervention may have unforeseen consequences, and requires testing as with any other new technology [32]. The digital media and paper version had similar effectiveness for affecting medicine use, but this study provides emerging evidence that digital intervention may be superior for services that require referral.

Digital delivery changes how participants interact with and experience the intervention. Integration with the patient electronic health record reduces the effort required to create new patient requests, such as actively inviting patients to a follow-up appointment. Therefore, the incorporation of the intervention in a clinician's workflow may explain the increased number of GP visits and psychologist referrals after the digital intervention when compared to the usual intervention. Creating a request to follow up a patient is easy to implement and unlikely to cause significant disruption. In contrast, reducing the dose of gabapentinoids, commonly indicated for pain, requires careful consideration and close patient contact and participation. Our results suggest that both postal and digital interventions are effective in promoting dose change, but it is possible that the digital medium advantage lies in creating triggers that can be easily followed.

The timing of this study is an important limitation of this study, as intervention delivery coincided with the initial restrictions implemented in Australia in response to the COVID-19 pandemic in March 2020. The week of the intervention, several policies to restrict gatherings and reduce risk of contagion were enacted [33], which influenced some of the metrics used in this study. Medicine dispensing was likely affected, with stockpiling occurring and a temporary lack of access. Additionally, many clinics were closed to avoid waiting room risks, and appointments via telehealth were funded by the Department of Health. This may have influenced intervention effectiveness, as the opportunity to adjust therapy was reduced; however, it is unlikely to have affected the assessment of mode of delivery as both arms of the trial would have been equally affected by the COVID-19 restrictions. Postal mail services were fully maintained during restrictions.

This study also provides a foundation for further research aimed at improving the effectiveness of audit and feedback in public health digital interventions. The effect size of conventional and digital audit and feedback interventions is usually small [1,7], and a clear methodology to improve effect remains an open question. Effect may be influenced by factors related to the recipient (eg, GP), behavior, or content and delivery of the intervention [34]. Using digital media enables nationwide programs such as Veterans' MATES to contribute to such research by creating repeated interventions at lower cost, with greater speed and precision.

Conclusion

This study showed a digitally delivered professional behavior change intervention had comparable effectiveness to a postal intervention and superior efficacy for referral services. Given the logistical benefits of digital delivery in nationwide programs (cost, speed, and precision), the results encourage exploration of this mode in future interventions.

Acknowledgments

This work was funded by the Australian Government Department of Veterans' Affairs (DVA) as part of the Veterans' Medicines Advice and Therapeutics Education Services (Veterans' MATES) program. The DVA reviewed this manuscript before submission but played no role in the study design, study execution, analysis or interpretation of data, writing of the manuscript, or decision to submit the paper for publication.

Veterans' MATES is provided by the University of South Australia, Quality Use of Medicines and Pharmacy Research Centre, in association with Discipline of General Practice, The University of Adelaide; Discipline of Public Health, The University of Adelaide; Repatriation General Hospital, Daw Park; National Prescribing Service (NPS) – Better choices, Better health; Australian Medicines Handbook; and the Drug and Therapeutics Information Service.

EER is supported by the National Health and Medical Research Council (GNT 1110139).

Editorial Notice

This randomized study was not prospectively registered. The authors explained that the trial was focused exclusively on its effect on providers, rather than patients. This trial was a modification of a previous existing intervention running continuously since 2004 (4 modules each year), so there was no additional enrollment for this particular trial. To reduce the bias in analysis, all analytic procedures (for this and all other interventions) were prospectively defined and submitted to the Australian Government Department of Veterans' Affairs for evaluation and approval. However, there is no formal registration number, and the plan is not made publicly available due to confidentiality reasons. The editor granted an exception from ICMJE rules mandating prospective registration of randomized trials because the risk of bias appears low, and the trial is targeted at providers. However, readers are advised to carefully assess the validity of any potential explicit or implicit claims related to primary outcomes or effectiveness, as the lack of registration means that authors could change their outcome measures retrospectively.

Authors' Contributions

AQA conceived of, designed, and conducted the final analysis for this study and drafted the manuscript. AQA, JPC, VTL, and EER developed the protocol and study approach. GMK, ER, LMKE, and NLP were involved in the data analysis. EER conceived

of and designed the study, and critically revised the manuscript for important intellectual content. All authors made important contributions to the theoretical approach and interpretation of insights. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Opt-out letter.
[DOCX File , 14 KB - jmir_v24i1e33873_app1.docx]

Multimedia Appendix 2
General practitioner communication campaign to inform about digital intervention.
[DOCX File , 248 KB - jmir_v24i1e33873_app2.docx]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1291 KB - jmir_v24i1e33873_app3.pdf]

References

1. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, et al. Audit and feedback: effects on professional practice and healthcare outcomes. Cochrane Database Syst Rev 2012 Jan;6:CD000259. [doi: 10.1002/14651858.CD000259.pub3] [Medline: 22696318]
2. Landis-Lewis Z, Brehaut JC, Hochheiser H, Douglas GP, Jacobson RS. Computer-supported feedback message tailoring: theory-informed adaptation of clinical audit and feedback for learning and behavior change. Implement Sci 2015 Jan 21;10(1):12-12 [FREE Full text] [doi: 10.1186/s13012-014-0203-z] [Medline: 25603806]
3. Dombrowski SU, O'Carroll RE, Williams B. Form of delivery as a key 'active ingredient' in behaviour change interventions. Br J Health Psychol 2016 Nov;21(4):733-740. [doi: 10.1111/bjhp.12203] [Medline: 27709824]
4. Black N, Eisma MC, Viechtbauer W, Johnston M, West R, Hartmann-Boyce J, et al. Variability and effectiveness of comparator group interventions in smoking cessation trials: a systematic review and meta-analysis. Addiction 2020 Sep 11;115(9):1607-1617 [FREE Full text] [doi: 10.1111/add.14969] [Medline: 32043675]
5. Webb TL, Joseph J, Yardley L, Michie S. Using the internet to promote health behavior change: a systematic review and meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. J Med Internet Res 2010;12(1):e4 [FREE Full text] [doi: 10.2196/jmir.1376] [Medline: 20164043]
6. van Genugten L, Dusseldorp E, Webb TL, van EP. Which Combinations of Techniques and Modes of Delivery in Internet-Based Interventions Effectively Change Health Behavior? A Meta-Analysis. J Med Internet Res 2016;18(6):e155 [FREE Full text] [doi: 10.2196/jmir.4218] [Medline: 27268104]
7. Tuti T, Nzinga J, Njoroge M, Brown B, Peek N, English M, et al. A systematic review of electronic audit and feedback: intervention effectiveness and use of behaviour change theory. Implement Sci 2017 May 12;12(1):1-20 [FREE Full text] [doi: 10.1186/s13012-017-0590-z] [Medline: 28494799]
8. Grimshaw J, Ivers N, Linklater S, Foy R, Francis JJ, Gude WT, Audit and Feedback MetaLab. Reinvigorating stagnant science: implementation laboratories and a meta-laboratory to efficiently advance the science of audit and feedback. BMJ Qual Saf 2019 May 09;28(5):416-423 [FREE Full text] [doi: 10.1136/bmjqs-2018-008355] [Medline: 30852557]
9. Morrison EE, Sandilands EA, Webb DJ. Gabapentin and pregabalin: do the benefits outweigh the harms? J R Coll Physicians Edinb 2017 Dec;47(4):310-313. [doi: 10.4997/JRCPE.2017.402] [Medline: 29537399]
10. Goodman CW, Brett AS. Gabapentin and Pregabalin for Pain - Is Increased Prescribing a Cause for Concern? N Engl J Med 2017 Aug 03;377(5):411-414. [doi: 10.1056/NEJMp1704633] [Medline: 28767350]
11. Quality Use of Medicines and Pharmacy Research Centre. Veterans' MATES. 2021. URL: https://www.veteransmates.net.au/ [accessed 2021-12-29]
12. Bandura A. Human agency in social cognitive theory. Am Psychol 1989 Sep;44(9):1175-1184. [doi: 10.1037/0003-066x.44.9.1175] [Medline: 2782727]
13. Prochaska JO, Velicer WF, Rossi JS, Goldstein MG, Marcus BH, Rakowski W, et al. Stages of change and decisional balance for 12 problem behaviors. Health Psychol 1994 Jan;13(1):39-46. [doi: 10.1037//0278-6133.13.1.39] [Medline: 8168470]
14. Green L, Kreuter M, Green L. Health program planning: an educational and ecological approach. New York: McGraw-Hill; 2005.
15. Ivers NM, Grimshaw JM, Jamtvedt G, Flottorp S, O'Brien MA, French SD, et al. Growing literature, stagnant science? Systematic review, meta-regression and cumulative analysis of audit and feedback interventions in health care. J Gen Intern Med 2014 Nov 26;29(11):1534-1541 [FREE Full text] [doi: 10.1007/s11606-014-2913-y] [Medline: 24965281]

16. Andrade AQ, LeBlanc VT, Kalisch-Ellett LM, Pratt NL, Moffat A, Blacker N, et al. Determinants of usefulness in professional behaviour change interventions: observational study of a 15-year national program. BMJ Open 2020 Oct 14;10(10):e038016 [FREE Full text] [doi: 10.1136/bmjopen-2020-038016] [Medline: 33055116]
17. Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. Milbank Q 2004;82(4):581-629 [FREE Full text] [doi: 10.1111/j.0887-378X.2004.00325.x] [Medline: 15595944]
18. The Royal Australian College of General Practitioners. General Practice: Health of the Nation. Melbourne, Victoria: RACGP; 2019.
19. Jha AK, Doolan D, Grandt D, Scott T, Bates DW. The use of health information technology in seven nations. Int J Med Inform 2008 Dec;77(12):848-854. [doi: 10.1016/j.ijmedinf.2008.06.007] [Medline: 18657471]
20. Roughead EE, Kalisch Ellett LM, Ramsay EN, Pratt NL, Barratt JD, LeBlanc VT, et al. Bridging evidence-practice gaps: improving use of medicines in elderly Australian veterans. BMC Health Serv Res 2013 Dec 12;13:514 [FREE Full text] [doi: 10.1186/1472-6963-13-514] [Medline: 24330781]
21. Kalisch Ellett LM, Pratt NL, Sluggett JK, Ramsay EN, Kerr M, LeBlanc VT, et al. Sustaining practice change in health care: the impact of a national quality improvement program on the uptake of collaborative medicines reviews. J Pharm Pract Res 2018 Jun 15;48(3):222-230. [doi: 10.1002/jppr.1379]
22. Kalisch Ellett LM, Pratt NL, Sluggett JK, Ramsay EN, Kerr M, LeBlanc VT, et al. Patient-specific prescriber feedback can increase the rate of osteoporosis screening and treatment: results from two national interventions. Arch Osteoporos 2017 Dec;12(1):17. [doi: 10.1007/s11657-017-0309-4] [Medline: 28188561]
23. Pratt NL, Kalisch Ellett LM, Sluggett JK, Ramsay EN, Kerr M, LeBlanc VT, et al. Commitment questions targeting patients promotes uptake of under-used health services: Findings from a national quality improvement program in Australia. Soc Sci Med 2015 Nov;145:1-6. [doi: 10.1016/j.socscimed.2015.09.019] [Medline: 26432175]
24. Pratt NL, Kalisch Ellett LM, Sluggett JK, Gadzhanova SV, Ramsay EN, Kerr M, et al. Use of proton pump inhibitors among older Australians: national quality improvement programmes have led to sustained practice change. Int J Qual Health Care 2017 Feb 01;29(1):75-82. [doi: 10.1093/intqhc/mzw138] [Medline: 27920248]
25. Kalisch Ellett LM, Lim R, Pratt NL, Kerr M, Ramsay EN, LeBlanc TV, et al. Reducing hypnotic use in insomnia management among Australian veterans: results from repeated national interventions. BMC Health Serv Res 2018 Aug 09;18(1):626-636 [FREE Full text] [doi: 10.1186/s12913-018-3443-9] [Medline: 30092801]
26. Quality Use of Medicines and Pharmacy Research Centre. Recovering from pain: strategies that can help. Veterans' MATES. URL: https://www.veteransmates.net.au/topic-58-veterans-advice [accessed 2021-12-29]
27. Bergman U. The history of the Drug Utilization Research Group in Europe. Pharmacoepidemiol Drug Saf 2006 Feb;15(2):95-98. [doi: 10.1002/pds.1171] [Medline: 16329154]
28. Candlish J, Teare MD, Dimairo M, Flight L, Mandefield L, Walters SJ. Appropriate statistical methods for analysing partially nested randomised controlled trials with continuous outcomes: a simulation study. BMC Med Res Methodol 2018 Oct 11;18(1):105-122 [FREE Full text] [doi: 10.1186/s12874-018-0559-x] [Medline: 30314463]
29. Seabold S, Perktold J. Statsmodels: Econometric and Statistical Modeling with Python. In: Proceedings of the 9th Python in Science Conference. 2010 Presented at: 9th Python in Science Conference; June 28-July 3, 2010; Austin, TX p. 92-96 URL: http://conference.scipy.org/proceedings/scipy2010/pdfs/seabold.pdf
30. Davidson-Pilon C. lifelines: survival analysis in Python. JOSS 2019 Aug;4(40):1317-1311. [doi: 10.21105/joss.01317]
31. Odone A, Buttigieg S, Ricciardi W, Azzopardi-Muscat N, Staines A. Public health digitalization in Europe. Eur J Public Health 2019 Oct 01;29(Supplement_3):28-35 [FREE Full text] [doi: 10.1093/eurpub/ckz161] [Medline: 31738441]
32. McKee M, van Schalkwyk MCI, Stuckler D. The second information revolution: digitalization brings opportunities and concerns for public health. Eur J Public Health 2019 Oct 01;29(Supplement_3):3-6 [FREE Full text] [doi: 10.1093/eurpub/ckz160] [Medline: 31738440]
33. Desborough J, Hall Dykgraaf S, de Toca L, Davis S, Roberts L, Kelaher C, et al. Australia's national COVID-19 primary care response. Med J Aust 2020 Aug 04;213(3):104-106.e1 [FREE Full text] [doi: 10.5694/mja2.50693] [Medline: 32623740]
34. Colquhoun HL, Carroll K, Eva KW, Grimshaw JM, Ivers N, Michie S, et al. Advancing the literature on designing audit and feedback interventions: identifying theory-informed hypotheses. Implement Sci 2017 Sep 29;12(1):117-110 [FREE Full text] [doi: 10.1186/s13012-017-0646-0] [Medline: 28962632]

Abbreviations

DDD: defined daily dose
DVA: Department of Veterans' Affairs
GP: general practitioner
HL7: Health Level Seven
MATES: Medicines Advice and Therapeutics Education Services
RCT: randomized controlled trial
WHO: World Health Organization

Edited by R Kukafka; submitted 27.09.21; peer-reviewed by MS Shafi, S Pesälä; comments to author 13.11.21; revised version received 15.11.21; accepted 21.11.21; published 10.01.22.

Please cite as:
Andrade AQ, Calabretto JP, Pratt NL, Kalisch-Ellett LM, Kassie GM, LeBlanc VT, Ramsay E, Roughead EE
Implementation and Evaluation of a Digitally Enabled Precision Public Health Intervention to Reduce Inappropriate Gabapentinoid Prescription: Cluster Randomized Controlled Trial
J Med Internet Res 2022;24(1):e33873
URL: https://www.jmir.org/2022/1/e33873
doi:10.2196/33873
PMID:35006086

Original Paper

Improving Diabetes-Related Biomedical Literature Exploration in the Clinical Decision-making Process via Interactive Classification and Topic Discovery: Methodology Development Study

Adrian Ahne[1,2], MSc; Guy Fagherazzi[3], MSc, PhD; Xavier Tannier[4], Prof Dr; Thomas Czernichow[2], MSc, PhD; Francisco Orchard[2], BSc

[1]Exposome and Heredity team, Center of Epidemiology and Population Health, Hospital Gustave Roussy, Inserm, Paris-Saclay University, Villejuif, France

[2]Epiconcept Company, Paris, France

[3]Deep Digital Phenotyping Research Unit, Department of Population Health, Luxembourg Institute of Health, Luxembourg, Luxembourg

[4]Laboratoire d'Informatique Medicale et d'Ingenierie des Connaissances pour la e-Sante, Limics, Inserm, University Sorbonne Paris Nord, Sorbonne University, Paris, France

Corresponding Author:
Adrian Ahne, MSc
Exposome and Heredity team, Center of Epidemiology and Population Health, Hospital Gustave Roussy
Inserm
Paris-Saclay University
20 Rue du Dr Pinel
Villejuif, 94800
France
Phone: 33 142115386
Email: adrian.ahne@protonmail.com

Abstract

Background: The amount of available textual health data such as scientific and biomedical literature is constantly growing and becoming more and more challenging for health professionals to properly summarize those data and practice evidence-based clinical decision making. Moreover, the exploration of unstructured health text data is challenging for professionals without computer science knowledge due to limited time, resources, and skills. Current tools to explore text data lack ease of use, require high computational efforts, and incorporate domain knowledge and focus on topics of interest with difficulty.

Objective: We developed a methodology able to explore and target topics of interest via an interactive user interface for health professionals with limited computer science knowledge. We aim to reach near state-of-the-art performance while reducing memory consumption, increasing scalability, and minimizing user interaction effort to improve the clinical decision-making process. The performance was evaluated on diabetes-related abstracts from PubMed.

Methods: The methodology consists of 4 parts: (1) a novel interpretable hierarchical clustering of documents where each node is defined by headwords (words that best represent the documents in the node), (2) an efficient classification system to target topics, (3) minimized user interaction effort through active learning, and (4) a visual user interface. We evaluated our approach on 50,911 diabetes-related abstracts providing a hierarchical Medical Subject Headings (MeSH) structure, a unique identifier for a topic. Hierarchical clustering performance was compared against the implementation in the machine learning library scikit-learn. On a subset of 2000 randomly chosen diabetes abstracts, our active learning strategy was compared against 3 other strategies: random selection of training instances, uncertainty sampling that chooses instances about which the model is most uncertain, and an expected gradient length strategy based on convolutional neural networks (CNNs).

Results: For the hierarchical clustering performance, we achieved an F1 score of 0.73 compared to 0.76 achieved by scikit-learn. Concerning active learning performance, after 200 chosen training samples based on these strategies, the weighted F1 score of all MeSH codes resulted in a satisfying 0.62 F1 score using our approach, 0.61 using the uncertainty strategy, 0.63 using the CNN, and 0.45 using the random strategy. Moreover, our methodology showed a constant low memory use with increased number of documents.

Conclusions: We proposed an easy-to-use tool for health professionals with limited computer science knowledge who combine their domain knowledge with topic exploration and target specific topics of interest while improving transparency. Furthermore,

our approach is memory efficient and highly parallelizable, making it interesting for large Big Data sets. This approach can be used by health professionals to gain deep insights into biomedical literature to ultimately improve the evidence-based clinical decision making process.

(J Med Internet Res 2022;24(1):e27434) doi:10.2196/27434

KEYWORDS

evidence-based medicine; clinical decision making; clinical decision support; digital health; medical informatics; transparency; hierarchical clustering; active learning; classification; memory consumption; natural language processing

Introduction

Clinical Decision Support Systems for Literature Summary

Evidence-based medicine combines clinical experience with the value of the patient and the best available research information to guide decision making about clinical management [1]. In order for health care professionals to practice evidence-based medicine for clinical decision making properly, efficient literature search skills are necessary [2], yet limits in time, knowledge, or skills are frequent barriers [3], explaining why only 1 in every 5 medical decisions is based strictly on evidence [4]. Clinical decision support systems offer a possibility to assist health professionals in improving health care delivery by enhancing medical decisions with targeted clinical knowledge, patient information, and other health information [5]. However, major challenges for efficient clinical decision support are using clinical knowledge such as extracted free-text information and transforming it into a usable form and mining large clinical databases to create new clinical decision support [6]. High-quality clinical decision support capabilities for clinicians are needed to appropriately interpret the exponentially growing data [6,7], such as electronic health records, laboratory results, doctor-patient interactions, social media, and biomedical literature [8-11], to improve clinical knowledge in the decision process.

Machine Learning to Analyze Textual Data

Machine learning and in particular natural language processing (NLP) techniques offer a solution to transform these health data into actionable knowledge [12] such as disease phenotypes, patient cohort identification [13,14], and decision support [15].

Despite the progress of machine learning techniques, the adoption of these methods in real practice is limited when the models lack interpretability and explainability, which are essential in the health care domain [16,17], or when models are challenging to apply for people with limited computer science skills [18]. In addition, many of the existing machine learning approaches to biomedical data analysis do not make the effort to integrate available expert knowledge into their models to improve model interpretability [19].

Well-established methods to explore unstructured textual information are topic models, such as latent Dirichlet allocation [20], which connect documents that share similar patterns and discover patterns of word use. Alternatively, word embeddings such as Word2Vec [21,22], FastText [23], or Bidirectional Encoder Representations from Transformers (BERT) [24] can be combined with a clustering algorithm such as K-means [25] to cluster documents [9].

However, these algorithms suffer from several limitations. In most clustering algorithms, the number of topics to be determined must be defined beforehand [26]; topic models lack scalability, and applied on large corpora, they are memory intensive [27]. As these topics are synthetic, they do not take prior knowledge of humans regarding the corpus domain into consideration [27]. Furthermore, topic models and most clustering algorithms are static systems. It is not possible to add more documents with time to the model without a complete retraining. Last, these models are not interactive in the sense that a user can influence and act on the topic exploration.

Objectives

In this paper, we propose an online decision support algorithm that provides a way for nonexperts, people without computer or data science knowledge, to discover topics of interest and classify unstructured health text data. We propose a single methodology for biomedical document classification and topic discovery that improves interpretability, (2) we provide an open-source tool for users without programming skills that can run on machines with limited calculation power and on big data clusters, and (3) we evaluate this methodology on a real-world use case to show it can reach a near state-of-the-art performance when compared with noninteractive and noninterpretable systems.

With our methodology, we aim to analyze a wide set of different clinical texts in different scenarios. Scientific interest over time based on publications or the evolution of public health opinion in social media can be evaluated as our approach is dynamic in the sense that new documents can easily be added to the model allowing the analysis over time. Furthermore, the combination of free text and multiple-choice answers on surveys or extracting cohort participant opinions from free-text content such as questionnaires can be studied. Another use case will be the classification of medical-related documents such as medical records, reports, and patient feedback.

The aim of this study is not to set a new benchmark in terms of performance but rather to tackle the existing limitations of NLP approaches in terms of usability in the health care domain to ultimately improve the literature exploration in the clinical decision-making process.

Methods

High-Level Overview

In the proposed methodology, documents are clustered in a hierarchical tree in a top-down fashion. A user alters this tree in an iterative process via an interactive user interface until a user-defined clustering solution of the documents is obtained. A high-level overview of this process is shown in Figure 1. Two types of nodes, clustering and classification, exist in the tree. A clustering node splits documents in an unsupervised way based on automatically detected headwords best describing the overall documents that have passed through it. Classification nodes are placed at the top of the tree via user interaction as users discover topics being represented by positive and negative training instances, discovered by exploration, describing a user-defined concept. A classification node is a binary machine learning algorithm acting as a barrier that lets documents pass to the underlying nodes only if they correspond to the defined concept.

In step 0, all documents start at the root node, referred to as "In Scope." The documents are then streamed one by one to construct the tree from the top to the bottom. The initial built tree consists of the root node, and all underlying nodes are clustering nodes. This fully automatic hierarchical procedure is detailed in the next section. Based on the clustering tree created, at each iteration a user starts exploring the tree and tries to identify a clustering node that summarizes a specific topic or concept via the interface, which provides information about the headwords and most important documents for each clustering node. When such a node is identified (eg, a node regrouping documents referring to type 2 diabetes), the user first creates a classifier node through the interface. The user then chooses sample documents that refer to type 2 diabetes (the positive instances) and sample documents that do not refer to type 2 diabetes (the negative instances). These instances will serve as training data for the underlying machine learning classifier of the classifier node. At the end of an iteration, the classifier nodes are trained and a new clustering tree is built, taking the trained classifiers into consideration. The idea is that each classifier groups together the documents corresponding to its user-defined concept or theme in the subtree below it. In this subtree, the documents continue to be clustered, allowing the exploration of subconcepts. At the next iteration, the user can explore the newly created tree, create new classifiers, choose training instances, and fix possible misclassifications via the interface. A sample iteration is shown in Figure 2, where a user identifies a cluster node referring to type 2 diabetes and creates a classifier node in the following iteration.

At each iteration, several classifier nodes can be created. Classifier nodes are always children of another classifier node near the top of the tree and start with a single clustering node child. With this active interaction between the user and the system, each iteration improves the performance of the classifier nodes, resulting in a better regrouping of similar documents and finally leading the model to converge toward a better user-defined solution. The results of this interactive process are a fine-tuned visualization tool for a given corpus or domain and a cascade of classifiers able to drive new documents to the most appropriate node of the tree.

Figure 3 illustrates a sample tree obtained after several iterations containing classifier nodes at the top of the tree and clustering nodes that continue to cluster documents. NLP methods were applied to represent documents and words. Word embeddings were used that transform each word into a vector representation [21,22]. A useful property of these word vector representations is that words similar in semantics are also close in this word vector space. Cosine similarity, a widely used metric in text analysis, was used as the distance measurement to decide whether 2 words were similar in semantics [28,29]. To determine if 2 documents were similar, the average over the word vectors of the documents were compared.

In the following sections, our approach is detailed in 4 parts: (1) a novel hierarchical clustering algorithm that processes documents in a streaming fashion; (2) user-defined classifiers to target topics; (3) a visual user interface through which the user explores the tree, annotates documents, and corrects misclassifications; and (4) a fully parallelizable interactive and iterative process leading to an accelerated convergence and minimized user annotation effort by combining the interpretable tree structure with active learning.

The methodology is implemented in the programming language Scala and the large-scale data processing framework Apache Spark. The word embeddings are streamed using Apache Lucene. The visual interface was created using the JavaScript language and the visualization library D3. The client server interaction is implemented using the open source toolkit Akka.

Figure 1. Overview of user interaction with the visual interface. SVM: support vector machine.

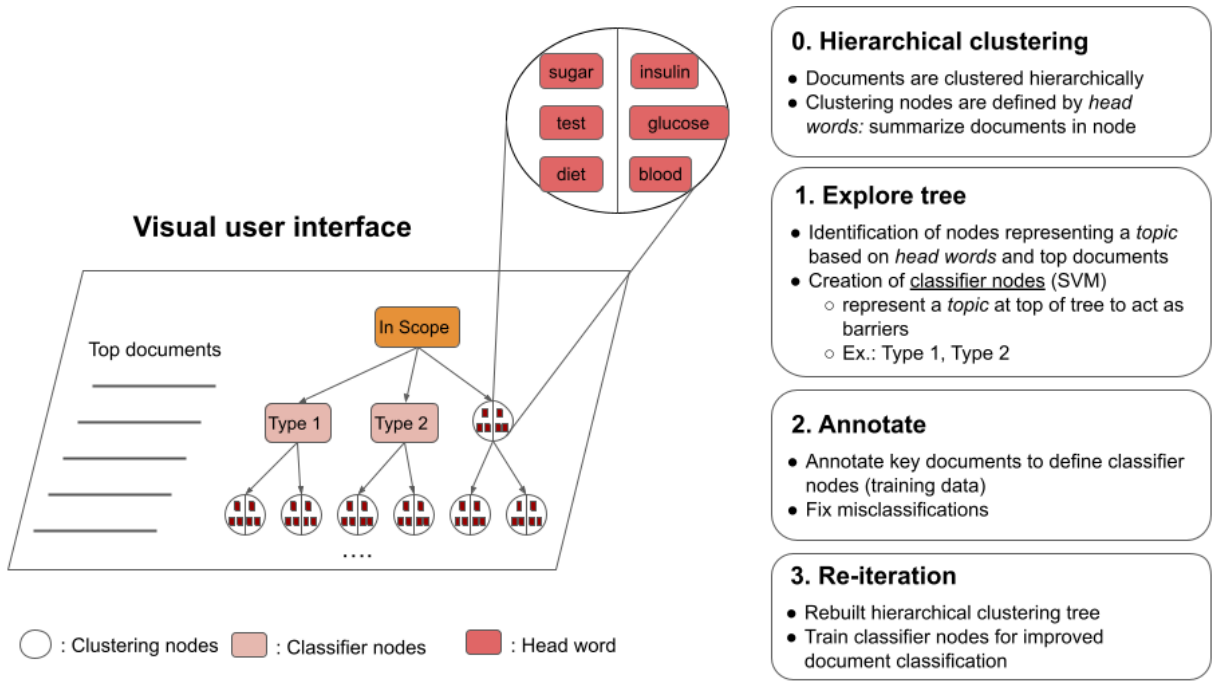


Figure 2. Iterative user interaction via the user interface following the 3 steps of exploring, annotating, and reiterating. To simplify, in iteration 1, no more classifiers are created. In a real-case scenario, a user usually defines several classifiers in the first iterations.

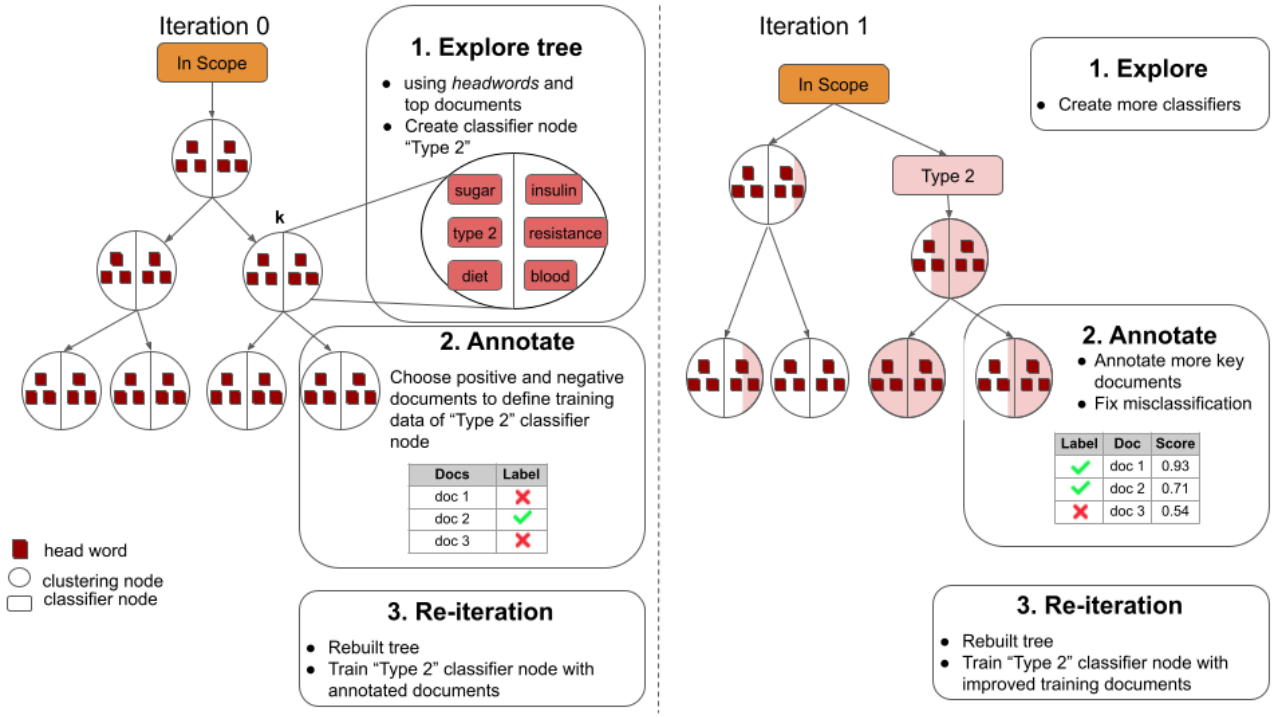
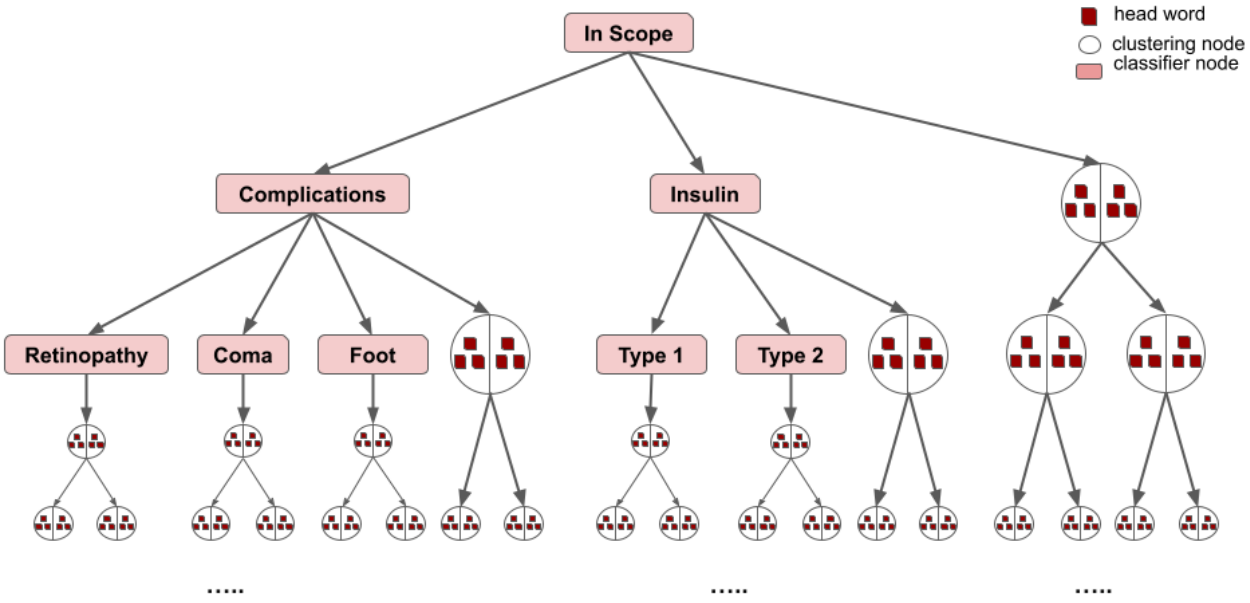


Figure 3. Classification and clustering tree after several iterations.



Hierarchical Clustering

Hierarchical clustering is a form of clustering in which the solution is presented in the form of trees. The different levels of the tree represent different levels of abstraction of the data. The consistency of the clustering solution at different levels of granularity allows flat partitions of different granularity to be extracted during data analysis, making them ideal for interactive exploration and visualization [30]. In many practical applications, it is more natural to discover the underlying structure of the data in a hierarchical manner rather than a flat one [31,32].

In our approach, the hierarchical clustering starts with a single clustering node that processes documents one by one leading to the creation of a binary tree structure where each node splits into two child nodes. During iterations, it is also possible to create several child nodes for a node through user interaction when classifier nodes are created. The tree is not equilibrated resulting in leaf nodes at different depths of the tree as some nodes stop splitting into children earlier than others.

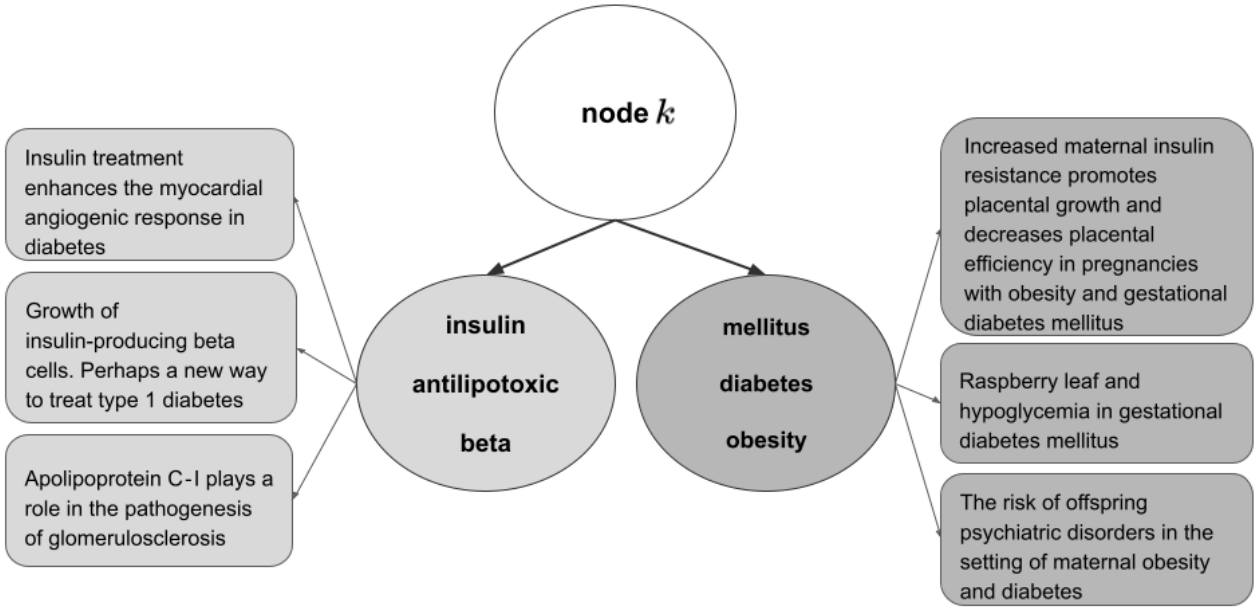
A key feature of our algorithm is that each document is processed individually, avoiding keeping all documents in memory or needing to know their total number, leading to a radical gain in memory use. This feature allows our approach to be dynamic, as more documents can be added over time allowing the study of cluster dynamics and evolution over time.

A clustering node is defined by headwords, which are the words that best represent the documents having descended the node. A clustering node can be split into further clustering nodes. Intuitively, the headwords of a node aim to summarize its documents. The objective is that a person could read the headwords and have an immediate understanding of the included documents, which considerably improves interpretability. We try to capture this notion by using the word embeddings semantic features and finding a set of tokens for which the sum of its word embeddings will be as close as possible to the sum of word embeddings of all tokens on all documents that went through the node. The semantic similarity of words is measured using cosine similarity.

To decide which path the document takes in the clustering process, given a document at a clustering node, it is compared to both clustering node children and associated to the one with the highest children score. This score is obtained by aggregating scores of each token based on the cosine similarity to its closest headword in the child nodes. For more information on the score calculation, please see Multimedia Appendix 1.

Each document traverses the tree and finds its way through comparison against the headwords of each node. If a document has reached a clustering node that is a tree leaf, two new clustering children are created and the document is then compared to the headwords to determine the child to which the document will be associated. Clustering node children will only be created when a minimum number of documents (default: 50) have passed the parent. The tree building continues until a user-defined number of maximal nodes is reached. After all documents have been processed to build the tree, the entire procedure is repeated, the documents are sent again one-by-one, such that headwords keep improving as long as the sum of all headword scores reaches a local maximum. Figure 4 provides an example of a real clustering node with its children and sample documents.

Figure 4. Real clustering example of a node, showing the headwords of its children. For each child, 3 sample titles of an abstract are provided. Note: only the titles and not the entire abstract are shown due to limited space.



Hierarchical Clustering Evaluation

As we use clustering as an exploration tool, our evaluation approach focuses on the overall quality of generated cluster hierarchies. One measure that takes into account the overall set of clusters represented in the hierarchical tree is the F1 score as introduced by Larsen and Aone [33] and used by Zhao and Karypis [30]. A detailed view of this score is provided in Multimedia Appendix 2.

Classification

A classifier node represents a user-defined topic. Internally, a support vector machine [34] classifier is embedded and predicts whether a document can be associated to the user-defined topic. Support vector machines have been shown to work well on textual data [35-37]. The classifier node acts as a filter and lets only these documents classified as the user-defined topic pass to the underlying nodes, where clustering continues.

The root node of each tree is a special “In Scope” classifier node. Using their domain knowledge, the user defines words that may represent what they are looking for and other words that may seem relevant. Assuming that a user expects to discover topics related to diabetes, possible words used as positive instances might be diabetes, insulin, hypoglycemia, pain, treatment, and risk. By default, stopwords such as and, of, or, and for are predefined as negative training examples. Based on the predefined words, the “In Scope” classifier is trained and used to separate locally relevant documents and noisy or irrelevant documents. Iteration 0 in Figure 2 illustrates the initial tree.

The user starts exploring the tree via the interface and tries to identify a clustering node that might represent a topic of interest based on headwords and most important documents. Targeting such a node *k* leads to the creation of a classifier node at the top of the tree, a clustering node child under the created classifier node, and a clustering node brother on the same level as the classification node as depicted in iteration 1 in Figure 2. A user chooses appropriate documents serving as positive and negative instances to train the classifier. When the tree is built again, each document entering the tree will first be fed to the type 2 classifier node. If the classifier predicts the document is related to type 2, the document passes the classifier node to its clustering node child. If the classifier rejects the document, the document is redirected to the clustering node brother, where clustering continues.

Iteration 1 in Figure 2 shows the purity of some nodes with regard to the proportion of documents related to type 2 diabetes in light red. Ideally, the nodes under a classifier only group documents relevant to the user-defined topic. In practice, and especially in the first iterations, this is not the case, as only a few instances served to train the classifier, affecting prediction performance. The user can interact with the interface to improve the classifier performance in 2 ways:

- Correcting misclassifications in the nodes under a classifier (by moving those documents to the negative training instances)
- Focusing on other parts of the tree that may contain documents related to type 2 that were not recognized by the classifier (to add them as positive training instances)

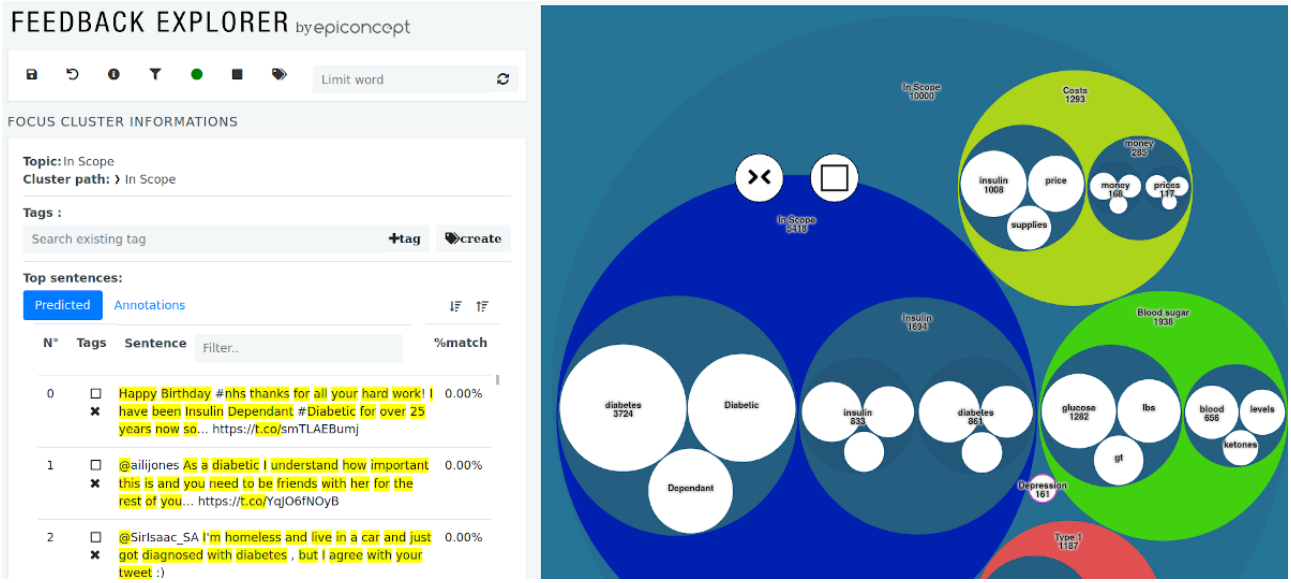
At the end of each iteration, the classifiers are retrained with the updated dataset, resulting in a steadily improving classification performance. During the exploration, if a user identifies a subtopic of an already created classifier, they can create a classifier child under a classifier node (Figure 3). In this iterative cycle, the user continues to create classifiers,

choose appropriate documents to train the classifiers, and correct misclassifications. This process eventually converges to form a user-desired clustering solution of topics of interest. A classifier node can increase its training set by using training instances of its surrounding classifiers. For more details, please see Multimedia Appendix 3.

Interactive Interface

An interactive interface has been developed in D3, jQuery, and JavaScript that visualizes the hierarchical clustering tree via nested circles (Figure 5). Moreover, each node provides information about its headwords and lists the sentences that run through this node ordered from the most to the least representative document. This order can also be reversed. On the bottom left, the documents for each node are shown. The colored nodes represent classifier nodes. Through the visualization, the transparency and interpretability of our methodology will be improved.

Figure 5. Visual user interface where colored circles represent user-defined topics (classifiers). Clicking on one of the nodes zooms into the node and shows the documents of the node on the bottom left. The headwords are shown in the white circles for each node.



Active Learning

Manual annotation is critical for the development and evaluation of machine learning classifiers to target topics. However, it is also time-consuming and expensive and thus remains challenging for research groups [38,39]. Active learning is a sample selection approach in the machine learning field that aims to minimize the annotation cost while maximizing the performance of machine learning–based models by choosing the training data in a smart way [40]. In active learning, only the most informative instances from an unlabeled dataset are selected to be labeled by an oracle (ie, a human). By choosing which instances should be labeled, an active learning algorithm can reduce time, effort, and resources needed to train a predictive model. This approach is attractive in scenarios where unlabeled data are widely available but labels are expensive. Several strategies exist to evaluate the informativeness of unlabeled data and choose training data [40]. Simplest and most commonly used is uncertainty sampling, in which the active learner chooses the instance about which it is the least certain how to label [41]. For example, for a binary probabilistic classifier, uncertainty sampling queries the instances where the posterior probability of being positive is nearest to 0.5. Other strategies used less often are the more theoretically motivated query-by-committee strategy [42] and the decision-theoretic approach in which the model selects the instance that would impart the greatest change to the current model if its label were known [43]. Active learning has been applied widely to textual data [35,44,45] and in clinical NLP [39,46]. Lu et al [47] showed that using modern word embeddings (Word2Vec, FastText, BERT) achieves significant improvement over more commonly used vector representations such as bag of words.

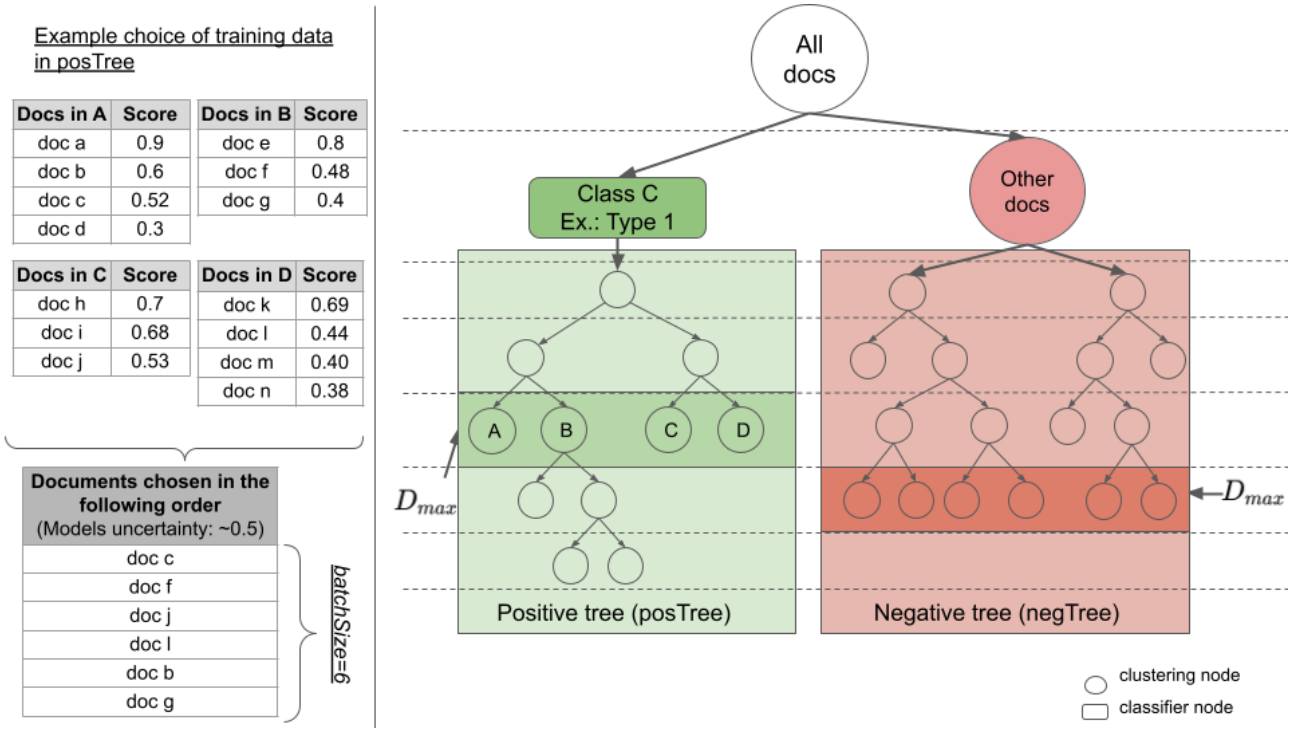
In this paper, we explore how our approach benefits from the combination of the active learning strategy uncertainty sampling and the hierarchical tree structure to minimize the user annotation effort and rapidly converge toward a user-guided clustering solution.

We developed an active learning strategy to automatically choose the best training instances for a given class, a Medical Subject Headings (MeSH) code in our case, by selecting documents from deeper levels of the tree. Figure 6 illustrates the details of the strategy for the MeSH code type 1 diabetes. In the first iteration, the tree is built containing only clustering nodes and followed by the creation of the type 1 classifier, which has no training instances yet. The depth level D_{max}, which is the level containing the most nodes in the tree, is then determined, and from each of those nodes, documents are chosen randomly consecutively. The parameter batchSize defines the number of chosen documents per iteration (default: 50). The first 50 documents will serve as initial instances to train the classifier node for the next iteration. In the next iteration, the tree is rebuilt taking the classifier with its instances into consideration leading to the tree in Figure 6. The tree can be separated into a positive tree (the subtree under the classifier, which concentrates documents of a specific topic: in this

example, type 1) and a negative tree (under the clustering brother of the classifier, which concentrates documents that don't refer to the class type 1). Usually the negative tree is larger as it concentrates documents from all other classes (MeSH codes). The idea is to add 25 (batchSize/2) documents as new training instances from each of the two subtrees. Similar to the first iteration, in both subtrees, the level D_{max} is identified and documents are chosen following the uncertainty sampling strategy contrary to the random selection of documents. From each node, the document the model is the most uncertain about to be of the specific class C is selected and added as a training instance. The assumption behind picking training instances from the level containing the most nodes is that documents in those nodes are best distributed in the vector space, which consequently provides well-distributed instances and avoids instances of being too similar. A concrete selection process on the positive tree is shown in Figure 6. Level D_{max} contains the nodes A, B, C, and D. In this example, the documents are chosen consecutively from these 4 nodes, which the model is most uncertain about. Uncertainty is measured as a prediction probability of being closest to 0.5.

The user has the choice of applying the automatic active learning strategy or the manual uncertainty sampling active learning strategy via the interface. In the interface, each node shows the headwords and documents in the node. The documents can be ordered from highest to lowest (to determine which documents are the most representative of the node) or lowest to highest (to determine the documents about which the model is most uncertain); the user can subsequently choose training instances based on these documents.

Figure 6. In active learning strategy, the positive tree is the subtree under the classifier node type 1, and the negative tree is the subtree under its clustering brother. On the left side, a sample of the document selection process is provided.



Active Learning Evaluation

Performance is addressed for each MeSH code individually. Given a MeSH code, all associated documents are considered the positive class while all other documents are considered the negative class. This leads to highly imbalanced datasets for most MeSH codes. Thus, it is also interesting to inspect the number of positive instances each strategy is able to detect.

A random subset of 2000 documents is chosen and randomly split into a training and test set of 1000 abstracts each. We evaluated the performance for 50, 100, 150, and 200 training instances per strategy to see if an increased performance can be observed in the first iterations. In the literature, most proposed active learning methods evaluated their performance only on a single measure, accuracy. However, Ramirez-Loaiza et al [48] showed that choosing only one metric to measure active learning performance can lead to unexpected and unwarranted conclusions. Hence, we evaluated our active learning method on accuracy, precision, recall, and F1 score.

The proposed methodology is embedded in an open source tool called Feedback Explorer (MadCap Software Inc). A video illustration of how Feedback Explorer functions is provided in a short video in Multimedia Appendix 4.

Results

Overview

In this section, we compare our hierarchical clustering and our active learning algorithm to the most popular existing algorithms. To that aim, we use a labeled classification dataset to assess the quality of our outcomes. The purpose of this study is not to establish a new state of the art but rather to show that our algorithm reaches near state-of-the-art performance while addressing the above-mentioned limitations of current systems

such as usability for nonexperts, memory consumption, and lack of interpretability.

Data

PubMed abstracts were downloaded from the US National Library of Medicine to test our algorithm [49]. In this corpus, abstracts are already classified in a hierarchical manner via MeSH codes [50]. We focused only on diabetes abstracts. Each selected abstract contained at least one diabetes MeSH code, which is an identifier for a topic. Due to a memory limitation of 30 GB for our analyses, we further reduced the dataset to be able to compare against more memory-intensive algorithms. To establish the maximum number of abstracts our system could handle, we started by setting the threshold at 1000, indicating the maximum number of abstracts per MeSH code. MeSH codes with fewer abstracts than the given threshold were fully included; otherwise a random sample of 1000 abstracts was chosen. We steadily increased this threshold by 1000 abstracts each iteration and reached a maximum threshold that our system could handle of 5000 abstracts per MeSH code. Table 1 provides an overview over all MeSH codes and the number of documents included for each code. The abstract publication dates range from 1949 to 2020.

In order to transform words into vectors, we used the biomedical word embeddings trained on biomedical texts from MEDLINE/PubMed [51], which are well adapted to our use case.

Table 1. Diabetes related MeSH[a] codes with number of documents per MeSH code.

Diabetes mellitus (C19.246)	N
Diabetes complications (C19.246.099)	5000
Diabetic angiopathies (C19.246.099.500)	3026
Diabetic foot (C19.246.099.500.191)	4424
Diabetic retinopathy (C19.246.099.500.382)	5000
Diabetic cardiomyopathies (C19.246.099.625)	386
Diabetic coma (C19.246.099.750)	97
Hyperglycemic hyperosmolar nonketotic coma (C19.246.099.750.490)	97
Diabetic ketoacidosis (C19.246.099.812)	1308
Diabetic nephropathies (C19.246.099.875)	5000
Diabetic neuropathies (C19.246.099.937)	3662
Diabetic foot (C19.246.099.937.250)	4424
Fetal macrosomia (C19.246.099.968)	1282
Diabetes, gestational (C19.246.200)	5000
Diabetes mellitus, experimental (C19.246.240)	5000
Diabetes mellitus, type 1 (C19.246.267)	5000
Wolfram syndrome (C19.246.267.960)	228
Diabetes mellitus, type 2 (C19.246.300)	5000
Diabetes mellitus, lipoatrophic (C19.246.300.500)	85
Donohue syndrome (C19.246.537)	39
Latent autoimmune diabetes in adults (C19.246.656)	16
Prediabetic state (C19.246.774)	1261

[a]MeSH: Medical Subject Headings

Hierarchical Clustering

We compared the hierarchical clustering part of Feedback Explorer with the hierarchical agglomerative clustering (HAC) algorithm. This algorithm has been implemented in several open-source libraries; we used the implementation in the popular machine learning library scikit-learn with complete linkage criterion, which provides an efficient hierarchical clustering algorithm [52].

For an equal comparison we ran both algorithms with two configurations, one with 32 leaf nodes and one with 64. We ran Feedback Explorer's clustering 10 times with random document order due to its streaming character which leads to different clustering solutions for a different order of documents. The F1 scores for the HAC algorithm were 0.76 for the 32 leaf nodes and 0.77 for the 64 leaf nodes, whereas the F1 scores for the Feedback Explorer clustering were 0.73 (95% CI 0.712-0.757) for the 32 leaf nodes and 0.74 (95% CI 0.717-0.760) for the 64 leaf nodes. Confidence intervals are not needed for the HAC algorithm as it is stable. In both cases, the HAC performance was superior; nevertheless, the F1 score for our approach with 0.73 and 0.74 comes close to the HAC performance.

Active Learning

To address the active learning classification performance, we compared 4 strategies. The first was the random strategy, in which the algorithm chose the documents randomly to train the classifier, followed by the uncertainty sampling strategy, in which the model chose the instances about which it was most uncertain [41]. Third was the Feedback Explorer strategy, introduced earlier in Methods. The fourth and last strategy, introduced by Zhang et al [45], combined convolutional neural networks [53] with the active learning strategy expected gradient length to classify text. Their proposed strategy selected documents if they contained words that were likely to most affect the word embeddings. This was achieved by calculating the expected gradient length with respect to the embeddings for each word [54]. The code of this approach is provided on GitHub by the authors [55].

Table 2 provides a performance overview over all MeSH codes (weighted average of accuracy, precision, recall, F1 score). The average confusion matrices over all MeSH codes for each strategy can be found in Multimedia Appendix 5. The scores after 200 training instances are similar within the 3 nonrandom approaches.

However, these averaged values mask the important variations of these systems depending on the MeSH codes they consider. In particular, MeSH codes with only a few relevant documents generally lead to very low performance. For a detailed overview of all MeSH codes, please refer to the table in Multimedia Appendix 6. For some MeSH codes, Feedback Explorer's strategy shows the highest performance after 200 iterations while for others the methods by Zhang et al [45] is superior. However, both strategies are similar in most cases. Multimedia Appendix 7 highlights specific results for 3 MeSH codes and additionally shows information about the positive and negative number of instances in the training set. For the MeSH code diabetes complications (D048909), Feedback Explorer reaches the highest performance after 200 training instances; for the MeSH code diabetic angiopathies (D003925), the method by Zhang et al [45] achieved best performance. The last MeSH code, diabetic cardiomyopathies (D058065), shows bad results for all strategies as only very few positive documents are contained in the dataset.

Table 2. Weighted average of active learning performance over all Medical Subject Headings codes.

# training data	Random				Uncertainty sampling				Feedback Explorer				CNN[a] Zhang			
	Acc[b]	Prec[c]	Rec[d]	F1[e]	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1	Acc	Prec	Rec	F1
50	0.87	0.62	0.57	0.51	0.83	0.56	0.60	0.50	0.88	0.63	0.44	0.49	0.81	0.24	0.31	0.20
100	0.86	0.62	0.51	0.49	0.88	0.68	0.64	0.62	0.90	0.71	0.51	0.56	0.86	0.39	0.59	0.42
150	0.88	0.68	0.46	0.47	0.90	0.75	0.62	0.63	0.90	0.75	0.59	0.60	0.88	0.52	0.72	0.55
200	0.89	0.62	0.43	0.45	0.91	0.77	0.53	0.61	0.91	0.71	0.58	0.62	0.90	0.58	0.79	0.63

[a]CNN: convolutional neural network.

[b]Acc: accuracy.

[c]Prec: precision.

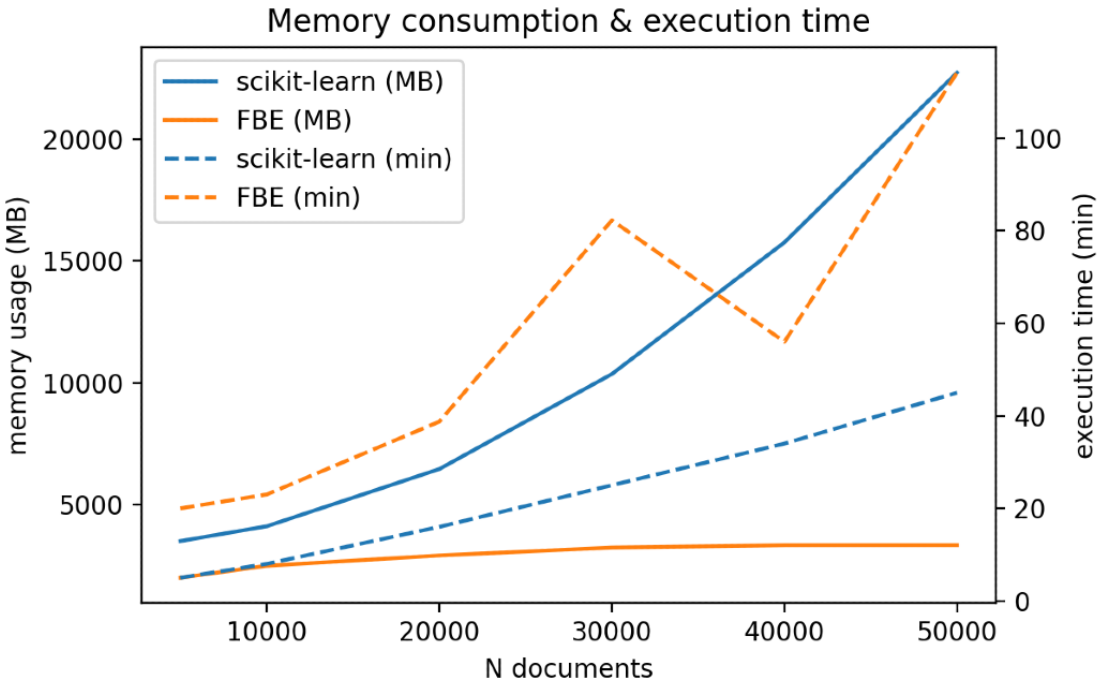
[d]Rec: recall.

[e]F1: F1 score.

Memory Consumption

Figure 7 provides an overview of the memory consumption in MB and execution time in minutes. Increasing the number of documents hardly changes the memory consumption for Feedback Explorer whereas HAC memory use grows exponentially. The memory efficiency of Feedback Explorer goes along with an expanding running time compared with the scikit-learn algorithm.

Figure 7. Memory consumption and execution times per volume of documents.



Discussion

Principal Findings

A visual interactive user interface has been developed, enabling users without computer science knowledge to discover and target topics of interest in unstructured clinical text to improve the literature exploration in the evidence-based clinical decision making process. An underlying HAC algorithm structures documents in an interpretable manner via headwords. The proposed method minimizes training instances effort in 2 ways: active learning strategy combines uncertainty sampling with the tree structure and manual intervention via the interface selects relevant documents about which the model is most uncertain as instances.

Feedback Explorer reaches near state-of-the-art performance in terms of hierarchical clustering as well as the active learning strategy. Furthermore, it addresses several existing limitations in common machine learning algorithms to extract information from text data: the challenge of adding domain knowledge to the model, the need to specify the desired number of clusters beforehand, the combination of classification and clustering in one methodology, and the difficulty of applying advanced machine learning algorithms for nonexperts without programming skills. These features make it an ideal asset for health professionals to analyze electronic health records, laboratory results, and social media data. We have shown that the memory consumption remains stable with an increased number of documents, which makes the algorithm particularly attractive in handling large datasets. The growing execution time can be minimized by heavier parallelization of the underlying Spark framework.

This methodology can be especially useful in complex clinical cases or for specialists who need to get a rapid overview of the existing literature concerning a specific topic.

Comparison With Prior Work

Several general purpose NLP systems have been developed to extract information from clinical text. The most frequently used tools are the Clinical Text Analysis and Knowledge Extraction System [56], MetaMap [57], and the Medical Language Extraction and Encoding System [58], according to the review by Wang et al [59]. These systems have been applied to information extraction tasks such as the identification of respiratory findings [60] or the detection of smoking status [56]. However, Zheng et al [61] showed that these systems are challenging to set up and customize, leading to general dissatisfaction that prevents adoptability.

The NLP Clinical Language Annotation, Modeling, and Processing toolkit (University of Texas Health Science Center at Houston) addresses this problem of difficult customization by also providing interaction via an interface to allow nonexperts to quickly develop customized clinical information extraction pipelines [62]. Besides the fact that the targeted task is quite different, this tool lacks generalizability beyond the domains it was trained on, and it is still difficult to add domain knowledge as opposed to our approach in which a user can use their expertise to specifically discover topics of interest [63].

In a recent literature survey concerning artificial intelligence in clinical decision support, Montani et al [64] emphasize the need for transparency and explainability in artificial intelligence systems such that users fully understand all generated suggestions. This is in line with our methodology as the user is directly involved and creates a user-defined solution. A more original approach is Plutchik, a voice-enabled, embodied artificial intelligence chatbot that can perform searches in

medical databases and retrieve and communicate medical information. But the integration of more sophisticated analysis methods, such as machine learning and deep learning methods, is still under development [65].

To the best of our knowledge, Feedback Explorer is the first decision support tool that combines topic exploration, topic targeting, user-friendly interface, minimization of memory consumption, and an annotation effort in a single methodology. This allows health professionals to rapidly gain insights about a clinical textual dataset to improve decision making.

Strength and Limitations

One of the key strengths of our methodology is that nonexperts with no programming knowledge are able to explore and target topics of interest in an unstructured textual dataset via an interactive and user-friendly interface. The fact that we visualize the headwords and the tree structure greatly improves transparency. Vellido Alcacena et al [66] also suggest that proper visualization can increase the transparency of machine learning. Moreover, since humans are directly included in the model creation, human interpretability is increased, as has also been shown by Lage et al [67]. Transparency of clinical decision support systems is key to ensure adoption by clinicians [68]. Due to its streaming nature, it is very memory efficient and can be used on a computer with limited memory. Additionally, the implementation is built on the basis of the large-scale data processing framework Apache Spark, which allows fast execution time through heavy parallelization of our algorithm resulting in the ability to handle large datasets. This is particularly interesting for the analysis of large text corpora, which usually are quite computation intensive [69]. Being able to dig into topics when an interesting cluster is found in combination with an interpretable result in terms of headwords and most important documents makes it particularly interesting for health care professionals. In addition, the proposed active learning strategy allows minimizing the annotation effort to train the classifiers by picking the most impactful training instances and enabling misclassification correction. The limitation in a classic clustering algorithm of specifying the desired number of clusters beforehand is addressed, as this parameter is not needed in our methodology. Currently it is still challenging to combine domain knowledge with topic extraction. Here, a health professional can apply their domain knowledge to search for specific topics of interest and test hypotheses to improve clinical decision making. This can be particularly helpful in the field of rare diseases, where clinical practice based on valid evidence is challenging [70]. Additionally, our model can be adapted to different languages by providing the corresponding word embeddings, which can be found easily in the web.

A limitation of our approach is that the number of classifiers a user can create is limited, as manual interaction is needed. In further investigations, our results should be confirmed on other datasets to ensure generalization and portability in other contexts. Also, the algorithm may construct marginally different tree structures that could affect data interpretation. The fact that the active learning performance is not always steadily increasing with more training instances but may sometimes oscillate is an open question in the active learning field [71]. This could be a future topic of investigation. A next step will be the evaluation of the proposed methodology on a sample of end users of various profiles and levels of expertise in clustering techniques. This will be the subject of a follow-up publication.

Conclusion

In this study, we proposed an interactive user interface for people without computer or data science knowledge to explore unstructured clinical text information as clinical decision support. The visualization of headwords and active participation of the user to drive the algorithm to converge to a user-defined solution greatly improves transparency. It combines several advantages such as using domain knowledge to target topics of interest, minimizing the manual annotation effort through active learning leading to a faster convergence, and minimizing memory consumption due to scalability, allowing processing of large corpora thanks to Spark's parallelism capabilities. We have shown that by combining all these advantages, we can reach near state-of-the-art performance. Such a tool can be of great assistance to health care professionals with limited computer science skills who want a rapid overview of specific topics while ultimately improving the literature exploration in the clinical decision-making process.

Authors' Contributions

AA and FO designed the study. AA collected the data. All authors discussed the methodology and evaluation strategy. AA and FO implemented the methodology and performed the evaluation analyses. All authors interpreted the results. AA drafted the manuscript. All authors commented on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Hierarchical clustering formulas.
[PDF File (Adobe PDF File), 166 KB - jmir_v24i1e27434_app1.pdf]

Multimedia Appendix 2
Hierarchical clustering evaluation.

[PDF File (Adobe PDF File), 51 KB - jmir_v24i1e27434_app2.pdf]

Multimedia Appendix 3
Increasing training set by using surrounding classifiers.
[PDF File (Adobe PDF File), 63 KB - jmir_v24i1e27434_app3.pdf]

Multimedia Appendix 4
Video illustration.
[MP4 File (MP4 Video), 103670 KB - jmir_v24i1e27434_app4.mp4]

Multimedia Appendix 5
Confusion matrices.
[PDF File (Adobe PDF File), 59 KB - jmir_v24i1e27434_app5.pdf]

Multimedia Appendix 6
Active learning performance for all Medical Subject Heading codes.
[PDF File (Adobe PDF File), 90 KB - jmir_v24i1e27434_app6.pdf]

Multimedia Appendix 7
Active learning performance for selected Medical Subject Heading codes for all 4 strategies.
[DOCX File , 20 KB - jmir_v24i1e27434_app7.docx]

References

1. Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. BMJ 1996 Jan 13;312(7023):71-72 [FREE Full text] [doi: 10.1136/bmj.312.7023.71] [Medline: 8555924]
2. Masic I, Miokovic M, Muhamedagic B. Evidence based medicine—new approaches and challenges. Acta Inform Med 2008;16(4):219-225 [FREE Full text] [doi: 10.5455/aim.2008.16.219-225] [Medline: 24109156]
3. van Dijk N, Hooft L, Wieringa-de Waard M. What are the barriers to residents practicing evidence-based medicine? A systematic review. Acad Med 2010;85(7):1163-1170. [doi: 10.1097/acm.0b013e3181d4152f]
4. Hernandez Medrano I, Tello Guijarro J, Belda C, Urena A, Salcedo I, Espinosa-Anke L, et al. Savana: re-using electronic health records with artificial intelligence. Int J Interact Multimed Artif Intell 2018;4(7):8. [doi: 10.9781/ijimai.2017.03.001]
5. Osheroff J. Improving Outcomes with Clinical Decision Support: An Implementer's Guide, 2nd Edition. London: Routledge; 2012.
6. Sittig DF, Wright A, Osheroff JA, Middleton B, Teich JM, Ash JS, et al. Grand challenges in clinical decision support. J Biomed Inform 2008 Apr;41(2):387-392 [FREE Full text] [doi: 10.1016/j.jbi.2007.09.003] [Medline: 18029232]
7. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. Lancet 2017 Jul;390(10092):415-423. [doi: 10.1016/s0140-6736(16)31592-6]
8. Althoff T, Clark K, Leskovec J. Large-scale analysis of counseling conversations: an application of natural language processing to mental health. Trans Assoc Comput Linguist 2016 Dec;4:463-476. [doi: 10.1162/tacl_a_00111]
9. Ahne A, Orchard F, Tannier X, Perchoux C, Balkau B, Pagoto S, et al. Insulin pricing and other major diabetes-related concerns in the USA: a study of 46,407 tweets between 2017 and 2019. BMJ Open Diabetes Res Care 2020 Jun 04;8(1):e001190 [FREE Full text] [doi: 10.1136/bmjdrc-2020-001190] [Medline: 32503810]
10. Gkotsis G, Oellrich A, Velupillai S, Liakata M, Hubbard TJP, Dobson RJB, et al. Characterisation of mental health conditions in social media using informed deep learning. Sci Rep 2017 Mar 22;7(1):45141 [FREE Full text] [doi: 10.1038/srep45141] [Medline: 28327593]
11. Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. Biomed Inform Insights 2016 Jan 19;8:BII.S31559. [doi: 10.4137/bii.s31559]
12. Biomedical text mining and its applications in cancer research. URL: https://reference.medscape.com/medline/abstract/23159498 [accessed 2020-12-02]
13. Gottesman O, Kuivaniemi H, Tromp G, Faucett WA, Li R, Manolio TA, et al. The Electronic Medical Records and Genomics (eMERGE) Network: past, present, and future. Genet Med 2013 Oct;15(10):761-771 [FREE Full text] [doi: 10.1038/gim.2013.72] [Medline: 23743551]
14. Xu H, Fu Z, Shah A, Chen Y, Peterson NB, Chen Q, et al. Extracting and integrating data from entire electronic health records for detecting colorectal cancer cases. AMIA Annu Symp Proc 2011;2011:1564-1572 [FREE Full text] [Medline: 22195222]
15. Demner-Fushman D, Chapman WW, McDonald CJ. What can natural language processing do for clinical decision support? J Biomed Inform 2009 Oct;42(5):760-772 [FREE Full text] [doi: 10.1016/j.jbi.2009.08.007] [Medline: 19683066]

16. Vellido A. The importance of interpretability and visualization in machine learning for applications in medicine and health care. Neural Comput Applic 2019 Feb 04;32(24):18069-18083. [doi: 10.1007/s00521-019-04051-w]
17. Ravi D, Wong C, Deligianni F, Berthelot M, Andreu-Perez J, Lo B, et al. Deep learning for health informatics. IEEE J Biomed Health Inform 2017 Dec;21(1):4-21. [doi: 10.1109/JBHI.2016.2636665] [Medline: 28055930]
18. Islam MS, Hasan MM, Wang X, Germack HD, Noor-E-Alam M. A systematic review on healthcare analytics: application and theoretical perspective of data mining. Healthcare (Basel) 2018 May 23;6(2):1 [FREE Full text] [doi: 10.3390/healthcare6020054] [Medline: 29882866]
19. Bhanot G, Biehl M, Villmann T, Zühlke D. Biomedical data analysis in translational research: integration of expert knowledge and interpretable models. 2017 Presented at: 25th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2017; 2017; Brussels p. 177-186 URL: https://www.rug.nl/research/portal/en/publications/biomedical-data-analysis-in-translational-research(0d1c2bbf-74ae-4132-9f02-a22e2f6af301).html
20. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res 2003;3:993-1022 [FREE Full text]
21. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. ArXiv. Preprint posted online on September 30, 2013. [FREE Full text]
22. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. ArXiv. Preprint posted online on October 30, 2013. [FREE Full text]
23. Bojanowski P, Grave E, Joulin A, Mikolov T. Enriching word vectors with subword information. ArXiv. Preprint posted online on June 30, 2017. [FREE Full text] [doi: 10.1162/tacl_a_00051]
24. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. ArXiv. Preprint posted online on May 30, 2019. [FREE Full text]
25. MacQueen J. Some methods for classification and analysis of multivariate observations. Proc 5th Berkeley Symp Math Stat Prob 1967:1 [FREE Full text]
26. Lee S, Baker J, Song J. An empirical comparison of four text mining methods. 43rd Hawaii Int Conf Syst Sci 2010:1-10. [doi: 10.1109/hicss.2010.48]
27. Ha-Thuc V, Srinivasan P. Topic models and a revisit of text-related applications. 2008 Presented at: 2nd PhD Workshop Inf Knowl Manag; 2008; Napa Valley p. 25. [doi: 10.1145/1458550.1458556]
28. Strehl E, Strehl J, Ghosh R. Impact of similarity measures on web-page clustering. Workshop Artif Intell Web Search 2000:58-64 [FREE Full text]
29. Tan P, Steinbach M, Karpatne A. Introduction to Data Mining, 2nd Edition. New York: Pearson; 2018.
30. Zhao Y, Karypis G. Evaluation of hierarchical clustering algorithms for document datasets. CIKM '02: Proc 11th Int Conf Inf Knowl Manag 2002:515-524. [doi: 10.21236/ada439551]
31. Sander J, Qin X, Lu Z. Automatic extraction of clusters from hierarchical clustering representations. In: Whang K, Jeon J, Shim K, editors. Advances in Knowledge Discovery and Data Mining, Volume 2637. Heidelberg: Springer Berlin; 2003:75-87.
32. Duda R, Hart P, Stork D. Pattern Classification, 2nd Edition. Hoboken: John Wiley & Sons, Inc; 2001.
33. Larsen B, Aone C. Fast and effective text mining using linear-time document clustering. Proc 5th ACM SIGKDD Int Conf Knowl Discov Data Mining 1999:16-22. [doi: 10.1145/312129.312186]
34. Vapnik V. Estimation of Dependences Based on Empirical Data. New York: Springer-Verlag; 2006.
35. Tong S, Koller D. Support vector machine active learning with applications to text classification. J Mach Learn Res 2002;2:45-66. [doi: 10.1162/153244302760185243]
36. Lo SL, Chiong R, Cornforth D. Using support vector machine ensembles for target audience classification on Twitter. PLoS One 2015 Apr 13;10(4):e0122855 [FREE Full text] [doi: 10.1371/journal.pone.0122855] [Medline: 25874768]
37. Marafino B, Davies J, Bardach N, Dean M, Dudley R. N-gram support vector machines for scalable procedure and diagnosis classification, with applications to clinical free text data from the intensive care unit. J Am Med Inform Assoc 2014;21(5):871-875 [FREE Full text] [doi: 10.1136/amiajnl-2014-002694] [Medline: 24786209]
38. Doğan RI, Leaman R, Lu Z. NCBI disease corpus: a resource for disease name recognition and concept normalization. J Biomed Inform 2014 Feb;47:1-10 [FREE Full text] [doi: 10.1016/j.jbi.2013.12.006] [Medline: 24393765]
39. Chen Y, Lasko TA, Mei Q, Denny JC, Xu H. A study of active learning methods for named entity recognition in clinical text. J Biomed Inform 2015 Dec;58:11-18. [doi: 10.1016/j.jbi.2015.09.010]
40. Settles B. Settles, Active Learning Literature Survey. Active learning literature survey. 2010. URL: http://burrsettles.com/pub/settles.activelearning.pdf?source=post_page [accessed 2020-12-19]
41. Lewis D, Gale W. A sequential algorithm for training text classifiers. Proc 17th Ann Int ACM SIGIR Conf Res Devel Inf Retrieval 1994 Aug:3-12. [doi: 10.1007/978-1-4471-2099-5_1]
42. Seung H, Opper M, Sompolinsky H. Query by committee. Proc 5th Ann Workshop Comput Learn Theory 1992 Jul:287-294. [doi: 10.1145/130385.130417]
43. Settles B, Craven M, Ray S. Multiple-instance active learning. 2008 Presented at: Advances in Neural Information Processing Systems (NeurIPS); 2007; - p. 1289-1296 URL: https://www.biostat.wisc.edu/~craven/papers/settles-nips07.pdf
44. McCallum K, Nigam K. Employing EM and pool-based active learning for text classification.: ICML; 1998. URL: http://www.kamalnigam.com/papers/emactive-icml98.pdf [accessed 2021-12-12]

45. Zhang Y, Lease M, Wallace B. Active discriminative text representation learning. ArXiv. Preprint posted online on December 31, 2016 [FREE Full text]
46. Figueroa RL, Zeng-Treitler Q, Ngo LH, Goryachev S, Wiechmann EP. Active learning for clinical text classification: is it better than random sampling? J Am Med Inform Assoc 2012 Sep 01;19(5):809-816 [FREE Full text] [doi: 10.1136/amiajnl-2011-000648] [Medline: 22707743]
47. Lu J, Henchion M, MacNamee B. Investigating the effectiveness of representations based on word-embeddings in active learning for labeling text datasets. ArXiv. Preprint posted online on October 31, 2019. [FREE Full text]
48. Ramirez-Loaiza ME, Sharma M, Kumar G, Bilgic M. Active learning: an empirical study of common baselines. Data Min Knowl Disc 2017 Mar;31(2):287-313. [doi: 10.1007/s10618-016-0469-7]
49. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2018 Jan;46(1):D8-D13 [FREE Full text] [doi: 10.1093/nar/gks1189] [Medline: 23193264]
50. Rampil J. The National Library of Medicine PubMed. Anesthesiology 1997 Nov;87(5):1268. [doi: 10.1097/00000542-199711000-00053]
51. McDonald R, Brokos G, Androutsopoulos I. Deep relevance ranking using enhanced document-query interactions. ArXiv. Preprint posted online on September 30, 2018. [FREE Full text] [doi: 10.18653/v1/d18-1211]
52. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12(85):2825-2830 [FREE Full text]
53. Goodfellow Y, Bengio Y, Courville A. Deep Learning. Cambridge: MIT Press; 2016.
54. Settles B, Craven M. An analysis of active learning strategies for sequence labeling tasks. 2008 Presented at: Conference on Empirical Methods in Natural Language Processing—EMNLP 08; 2008; Honolulu p. 1070 URL: https://dl.acm.org/doi/pdf/10.5555/1613715.1613855 [doi: 10.3115/1613715.1613855]
55. Active learning for neural networks. URL: https://github.com/yezhang-xiaofan/Active-Learning-for-Neural-Networks [accessed 2022-01-09]
56. Savova GK, Masanz JJ, Ogren PV, Zheng J, Sohn S, Kipper-Schuler KC, et al. Mayo clinical Text Analysis and Knowledge Extraction System (cTAKES): architecture, component evaluation and applications. J Am Med Inform Assoc 2010;17(5):507-513 [FREE Full text] [doi: 10.1136/jamia.2009.001560] [Medline: 20819853]
57. Aronson AR, Lang F. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc 2010;17(3):229-236 [FREE Full text] [doi: 10.1136/jamia.2009.002733] [Medline: 20442139]
58. Friedman C. Towards a comprehensive medical language processing system: methods and issues. Proc AMIA Annu Fall Symp 1997:595-599 [FREE Full text] [Medline: 9357695]
59. Wang Y, Wang L, Rastegar-Mojarad M, Moon S, Shen F, Afzal N, et al. Clinical information extraction applications: a literature review. J Biomed Inform 2018 Jan;77:34-49 [FREE Full text] [doi: 10.1016/j.jbi.2017.11.011] [Medline: 29162496]
60. Chapman WW, Fiszman M, Dowling JN, Chapman BE, Rindflesch TC. Identifying respiratory findings in emergency department reports for biosurveillance using MetaMap. Stud Health Technol Inform 2004;107(Pt 1):487-491. [Medline: 15360860]
61. Zheng K, Vydiswaran VV, Liu Y, Wang Y, Stubbs A, Uzuner ?, et al. Ease of adoption of clinical natural language processing software: an evaluation of five systems. J Biomed Inform 2015 Dec;58 Suppl:S189-S196 [FREE Full text] [doi: 10.1016/j.jbi.2015.07.008] [Medline: 26210361]
62. Soysal E, Wang J, Jiang M, Wu Y, Pakhomov S, Liu H, et al. CLAMP: a toolkit for efficiently building customized clinical natural language processing pipelines. J Am Med Inform Assoc 2018 Mar 01;25(3):331-336 [FREE Full text] [doi: 10.1093/jamia/ocx132] [Medline: 29186491]
63. Bompelli A. Comparing NLP systems to extract entities of eligibility criteria in dietary supplements clinical trials using NLP-ADAPT. In: Michalowski M, Moskovitch R, editors. Artificial Intelligence in Medicine. AIME 2020. Lecture Notes in Computer Science, vol 12299. Cham: Springer; 2020:67-77.
64. Montani S, Striani M. Artificial intelligence in clinical decision support: a focused literature survey. Yearb Med Inform 2019 Aug 16;28(1):120-127 [FREE Full text] [doi: 10.1055/s-0039-1677911] [Medline: 31419824]
65. Bohle S. "Plutchik": artificial intelligence chatbot for searching NCBI databases. J Med Libr Assoc 2018 Oct 04;106(4):501-503 [FREE Full text] [doi: 10.5195/jmla.2018.500] [Medline: 30271296]
66. Vellido Alcacena J, Martín J, Rossi F. Seeing is believing: the importance of visualization in real-world machine learning applications. 2011. URL: https://upcommons.upc.edu/handle/2117/20273 [accessed 2020-12-09]
67. Lage A, Ross B, Kim B. Human-in-the-loop interpretability prior. ArXiv. Preprint posted online on October 30, 2018. [FREE Full text]
68. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. JAMA 2018 Dec 04;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
69. Yang T, Chen Y, Emer J. A method to estimate the energy consumption of deep neural networks. 2017 51st Asilomar Conf Signals Syst Comput 2017 Oct:1916-1920. [doi: 10.1109/acssc.2017.8335698]
70. Rath A, Salamon V, Peixoto S, Hivert V, Laville M, Segrestin B, et al. A systematic literature review of evidence-based clinical practice for rare diseases: what are the perceived and real barriers for improving the evidence and how can they be overcome? Trials 2017 Nov 22;18(1):556 [FREE Full text] [doi: 10.1186/s13063-017-2287-7] [Medline: 29166947]

71. Chen Y, Mani S, Xu H. Applying active learning to assertion classification of concepts in clinical text. J Biomed Inform 2012 Apr;45(2):265-272 [FREE Full text] [doi: 10.1016/j.jbi.2011.11.003] [Medline: 22127105]

Abbreviations

BERT: Bidirectional Encoder Representations from Transformers
HAC: hierarchical agglomerative clustering
MeSH: Medical Subject Headings
NLP: natural language processing

Edited by R Kukafka; submitted 25.01.21; peer-reviewed by K Silva, SM Ayyoubzadeh, X Lyu; comments to author 01.03.21; revised version received 06.04.21; accepted 10.11.21; published 18.01.22.

Please cite as:
Ahne A, Fagherazzi G, Tannier X, Czernichow T, Orchard F
Improving Diabetes-Related Biomedical Literature Exploration in the Clinical Decision-making Process via Interactive Classification and Topic Discovery: Methodology Development Study
J Med Internet Res 2022;24(1):e27434
URL: https://www.jmir.org/2022/1/e27434
doi:10.2196/27434
PMID:35040795

Original Paper

A Clinical Decision Support System for Sleep Staging Tasks With Explanations From Artificial Intelligence: User-Centered Design and Evaluation Study

Jeonghwan Hwang[1*], BSc, MSc; Taeheon Lee[1*], BSc, MSc; Honggu Lee[1], BSc, MSc, PhD; Seonjeong Byun[2], MD

[1]Looxid Labs, Seoul, Republic of Korea

[2]Department of Neuropsychiatry, Uijeongbu St Mary's Hospital, College of Medicine, The Catholic University of Korea, Uijeongbu-si, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Seonjeong Byun, MD
Department of Neuropsychiatry
Uijeongbu St Mary's Hospital, College of Medicine
The Catholic University of Korea
271, Chenbo-ro
Uijeongbu-si, 11765
Republic of Korea
Phone: 82 31 820 3946
Email: sunjung.byun@gmail.com

Abstract

Background: Despite the unprecedented performance of deep learning algorithms in clinical domains, full reviews of algorithmic predictions by human experts remain mandatory. Under these circumstances, artificial intelligence (AI) models are primarily designed as clinical decision support systems (CDSSs). However, from the perspective of clinical practitioners, the lack of clinical interpretability and user-centered interfaces hinders the adoption of these AI systems in practice.

Objective: This study aims to develop an AI-based CDSS for assisting polysomnographic technicians in reviewing AI-predicted sleep staging results. This study proposed and evaluated a CDSS that provides clinically sound explanations for AI predictions in a user-centered manner.

Methods: Our study is based on a user-centered design framework for developing explanations in a CDSS that identifies why explanations are needed, what information should be contained in explanations, and how explanations can be provided in the CDSS. We conducted user interviews, user observation sessions, and an iterative design process to identify three key aspects for designing explanations in the CDSS. After constructing the CDSS, the tool was evaluated to investigate how the CDSS explanations helped technicians. We measured the accuracy of sleep staging and interrater reliability with macro-F1 and Cohen κ scores to assess quantitative improvements after our tool was adopted. We assessed qualitative improvements through participant interviews that established how participants perceived and used the tool.

Results: The user study revealed that technicians desire explanations that are relevant to key electroencephalogram (EEG) patterns for sleep staging when assessing the correctness of AI predictions. Here, technicians wanted explanations that could be used to evaluate whether the AI models properly locate and use these patterns during prediction. On the basis of this, information that is closely related to sleep EEG patterns was formulated for the AI models. In the iterative design phase, we developed a different visualization strategy for each pattern based on how technicians interpreted the EEG recordings with these patterns during their workflows. Our evaluation study on 9 polysomnographic technicians quantitatively and qualitatively investigated the helpfulness of the tool. For technicians with <5 years of work experience, their quantitative sleep staging performance improved significantly from 56.75 to 60.59 with a *P* value of .05. Qualitatively, participants reported that the information provided effectively supported them, and they could develop notable adoption strategies for the tool.

Conclusions: Our findings indicate that formulating clinical explanations for automated predictions using the information in the AI with a user-centered design process is an effective strategy for developing a CDSS for sleep staging.

(J Med Internet Res 2022;24(1):e28659) doi:10.2196/28659

KEYWORDS

sleep staging; clinical decision support; user-centered design; medical artificial intelligence

Introduction

Background

Polysomnography is a systematic process for collecting physiological parameters during sleep and is a diagnostic tool for evaluating various sleep disorders. Physiological recordings obtained from an electroencephalogram (EEG), electrooculogram (EOG), and electromyogram (EMG) were inspected by polysomnographic technicians to obtain important sleep parameters. Sleep staging is the process of identifying periodic changes in sleep stages. Typically, sleep stages are identified for every 30-second signal or epoch. On the basis of the American Academy of Sleep Medicine; wake status; 3 non–rapid eye movement (REM) stages, namely N1, N2, and N3; and REM stages were identified from polysomnographic recordings [1]. Sleep staging is an essential task in sleep medicine, as sleep patterns contain critical information for analyzing overnight polysomnography. To be specific, crucial sleep parameters, such as the distribution of sleep stages, were extracted from the sleep staging results. For example, the N1 stage, which is difficult to differentiate from the wake stages, is used to calculate the time to sleep onset and total sleep time parameters. The detection of REM stages affects the calculation of REM latency after sleep, which is another important sleep parameter. Furthermore, the physiological characteristics associated with each sleep stage have been investigated to diagnose several sleep disorders, such as obstructive sleep apnea, narcolepsy, and REM sleep behavior disorder [2,3]. However, in polysomnography, sleep staging is a time-consuming and costly process because every epoch in an overnight recording must be manually inspected. Several algorithms have been introduced to automate this time-consuming and costly task [4-6].

Artificial Intelligence–Based Clinical Decision Support Systems for Sleep Staging

Advances in deep learning techniques have led to the development of clinical Artificial Intelligence (AI) systems with diagnostic performance comparable with that of human clinicians [4,7-9]. These models have been introduced to automate time-consuming diagnoses and annotation procedures in clinical fields. However, the full automation of diagnostic processes, where algorithmic counterparts completely replace human clinicians, is presently not available owing to several challenges: the reliability of model predictions [10], clinical soundness of model behaviors [11], and social consensus on the replacement [12]. Similarly, in sleep medicine, several studies have introduced AI algorithms to automate time-consuming sleep staging tasks, but manual reviews of the results after automated prediction remain mandatory [13,14]. Under these circumstances, systems to assist polysomnographic technicians during the review process are in demand. For example, prior work in human–AI interaction conceptualized a framework in which ambiguous portions in polysomnographic recordings are selectively prioritized for manual inspection [15].

Despite an increasing number of deep learning studies for sleep staging [4,5], implementing an adoptable clinical decision support system (CDSS) for clinical practice remains a challenging task. First, regarding clinical knowledge, most deep learning–based systems lack explainable factors, but clinical staff members require clinically sound systems [10,13,16]. Thus, the CDSS should provide users with the necessary explanations. Second, the user interface of the AI system should be practical in clinical environments, where the time and resources of clinicians are constrained [10,17]. Therefore, a tool design that promotes readability and accessibility of the AI model from the viewpoint of clinical practitioners is indispensable for integrating AI-based decision-making into the workflow of human technicians [10,18]. The development of such CDSSs is crucial because these tools could alleviate these time-consuming and costly clinical tasks. Furthermore, proper algorithmic assistance can enhance the performance of clinical practitioners [19].

Study Objectives

In this study, we introduce an AI-based CDSS for assisting polysomnographic technicians when reviewing the AI-generated sleep staging results. Our objective is to correctly understand the information required from the CDSS and to develop the system in a user-centered manner. Through an extensive user study, we determined the features desired in a sleep staging AI system that could successfully support sleep technicians. We formulated the development process of a tool to assist clinical practitioners effectively.

Methods

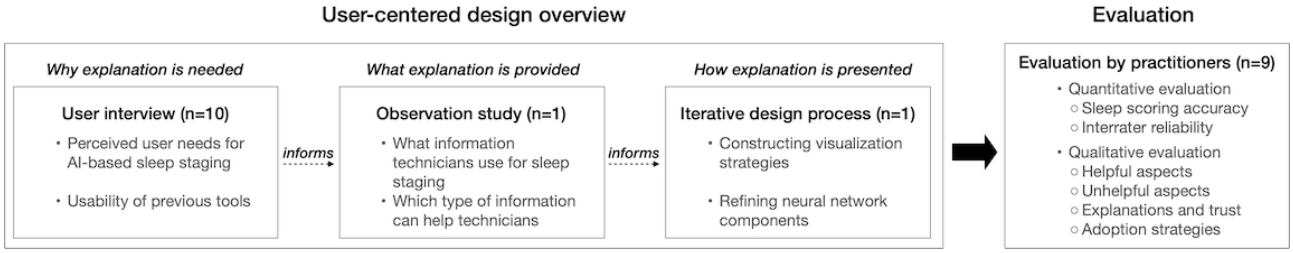
Study Design

This study aimed to understand what information should be provided to assist sleep technicians in collaborating with AI-based CDSS and to implement this system practically using a user-centered approach. Recent studies for designing explanations in CDSS propose frameworks that identify three key components from the perspectives of users: *why* information from CDSS is desired for a task, *what* content should be included in the explanation, and *how* explanations should be presented to users [20,21].

To define why users need explanations from the CDSS, the context within which users request explanations must be understood first. This question relates to the needs of the users and the purpose of the explanations. The perspective of users should determine the explanatory objective concerning the information that should be provided. A possible set of information that can be considered from this phase includes explanations for the input data, explanations related to the domain knowledge used in the task, causal information on how the system generates an output, and how results change with changes in input data [21,22]. Finally, several design factors, such as the units and format used for explanations, are considered when determining how information should be provided.

To design a CDSS within this framework, our development process included three phases: (1) interviews with polysomnographic technicians to identify why users might desire explanations from the CDSS when adopting AI-based sleep scoring systems, (2) user observations of how polysomnographic technicians score sleep stages from EEG recordings to determine the information that could help them, and (3) an iterative design process to construct a user-friendly CDSS interface that addresses the formulation of explanations in the system. After development, the polysomnographic technicians performed quantitative and qualitative evaluations of the system. In this section, we describe the objectives of each phase and explain how we conducted each phase (Figure 1).

Figure 1. Overall development process. AI: artificial intelligence.



Participants

Polysomnographic technicians with expertise in sleep staging were recruited for this study. Only technicians with a national license for medical laboratory technologists who were eligible to conduct polysomnography scoring were considered. To recruit participants with expertise in sleep scoring, we restricted their participation to those with experience in polysomnography scoring. We recruited 10 technicians to participate in the user interviews during the first phase and subsequent evaluation studies. We set the number of participants to 10, following previous studies on CDSSs, in which the number of participants was between 6 and 12 [15,23]. Among the technicians, we aimed to recruit 1 technician who could deeply engage in the development process by participating in the user observation and iterative design processes, which required regular meetings. We recruited technicians from secondary and tertiary hospitals rather than primary hospitals. Participants were recruited through emails sent to the polysomnographic technician community.

We recruited participants and divided them into two groups, *novice* technicians with <5 years of experience and *senior* technicians with >5 years of experience, to evaluate whether there were any differences in the helpfulness of the CDSS based on the amount of experience. On the basis of the Rasmussen skill-, rule-, and knowledge-based behavior model [24], we assumed that senior technicians would score stages subconsciously compared with novice technicians who consciously process the EEG characteristics. Here, we expected that novice technicians would more extensively refer to the provided explanation than senior technicians because novice technicians may find it difficult to quickly locate important EEG patterns. Thus, it was thought meaningful to investigate how our explanations affected technicians based on their skills.

Development Procedure

User Interview: Why Explanation Is Desired

We conducted user interviews with polysomnographic technicians to investigate why technicians would need explanations from the CDSS when AI-based support systems were adopted for sleep staging. During the interview, we first presented several questions regarding user needs during manual sleep staging and the perceptions of technicians regarding the utility of previous sleep staging AI tools. The technicians were asked whether they were using the automated sleep staging programs. Furthermore, the reasons for not adopting such automated sleep staging programs were investigated. Upon further investigation, we established the context in which explanations from AI were desired when reviewing automated sleep staging results. A user study was conducted using structured interviews with the sample questions listed in Textbox 1.

Textbox 1. Examples of interview questions in the user study.

Topic and question statement

User needs during manual sleep staging

1. How much time do you spend on a sleep staging task when performing polysomnography?
2. For sleep staging tasks, on which features of electroencephalogram recordings do you mainly focus?
3. Do you feel any need for assistance during sleep staging?

Utility of sleep staging artificial intelligence (AI) tools

1. There are several AI programs that automate sleep staging tasks; are you adopting them in your workflow? If not, what are the problems associated with these programs?
2. In which processes do you need AI programs to assist your sleep staging tasks?
3. Assuming that there is an AI program that automates sleep staging tasks and sleep technicians only need to review its scorings, in which context are explanations desired for an efficient review process?

User Observation: What Information Should Be Contained in Explanations

A user observation study was performed to understand the sleep staging conventions of clinical practitioners. From the observed sleep staging conventions, we aimed to construct a list of EEG characteristics to which technicians refer. During this study, hour-long weekly meetings were held over a month in which a participating technician scored EEG epochs in a think-aloud protocol. The technician was requested to verbally express how the information in the EEG recordings during sleep scoring was processed. Afterward, the technician reviewed the scoring with detailed explanations of the reasons for scoring the epochs with the annotated stages. The objective of these observation sessions was to formulate what information could assist technicians in reviewing predictions from AI algorithms. The observations were made based on characteristic EEG patterns such as sleep spindles, k-complexes, and frequency waves listed in the sleep manual [1]. We investigated how the listed EEG characteristics were inspected in practice. Subsequently, we grouped the EEG features into typical explanations that our CDSS could provide.

Iterative Design Process: How Explanations Can Be Presented

We conducted an iterative design process with a technician to identify how explanations should be presented to CDSS users. For 2 months, we held weekly 2-hour meetings.

The Template-Guided Neural Networks for Robust and Interpretable Sleep Stage Identification from EEG Recordings (TRIER) was selected as the AI algorithm for generating explanations. It is a convolutional neural network architecture used to process single-channel EEG data for sleep staging, and was proposed to extract clinically meaningful EEG wave shapes. This study demonstrated the possibility that features in the convolutional filters could be related to important EEG characteristics such as sleep spindles and k-complexes, with a sleep staging performance comparable with human raters with macro-F1 scores of 0.7-0.8 on public sleep data sets. We considered three components in the TRIER, namely convolutional filters, saliency values, and intermediate activation, as sources of information for generating explanations. These three components have been widely used in interpreting neural network operations in previous machine learning studies [25-30]. Detailed technical descriptions of these components are provided in Multimedia Appendix 1 [1,29-32].

During the iterations, we aimed to investigate whether the information contained in the above components could provide the desired information obtained from the user observation study. In these sessions, the technician inspected the features obtained by the neural network components and expressed an opinion on whether they could provide sufficient explanation for the task. Information from the components was refined based on the feedback. Subsequently, we chose the exact component for generating explanations from the neural network components. However, because the information in neural networks is numerical, adequate visualization is required to enhance the user-friendliness of the explanations. Therefore, we iteratively collected feedback on the representation format of the explanations during the later sessions. The technician tested the prototype versions of the proposed tool and provided feedback in terms of their intuitiveness and helpfulness. Consequently, visualization strategies were constructed for the explanations and overall interfaces.

Evaluation Study

Data Set Preparation

During the evaluation, technicians scored the sleep stages on sleep recordings from a public sleep EEG data set, the ISRUC-Sleep Dataset [33]. These data contain polysomnographic recordings obtained from 100 subjects with evidence of sleep disorders. This data set was collected from the Sleep Medicine Centre of the Hospital of Coimbra University. We adopted the public data set for sleep staging to calculate sleep staging performance based on the ground-truth labels provided in the data set. The characteristics of the data sets are summarized in Table 1.

The data were divided into a training set (80 participants), validation set (10 participants), and test set (10 participants). Only data samples from the training data set were used for training the deep learning models. We used the validation data set to select the model to be used for constructing the CDSS. The model with the best performance scores for the validation set was selected. The experimental results and corresponding findings were drawn exclusively from the test data set, which means that to avoid information leakage issues that may affect model accuracy, the data samples used for training the model were not used during the evaluation study.

To construct the data set for the evaluation study, we randomly extracted 15-minute EEG segments from the EEG recordings in the test data set. EEG segments with no changes in sleep stages were excluded from the selected segments. We evaluated the sleep scoring performance with 15-minute segments rather than whole-night polysomnography to evaluate the helpfulness of the tool effectively. Considering that technicians often skim through recordings and pay attention to EEG epochs with stage changes, the effectiveness of the system might not be revealed or hindered by the back-and-forth temporal relations between the sleep stages. This evaluation configuration was also adopted in a previous CDSS study for sleep staging [15]. In addition to the test set of 15-minute segments, we constructed a test data set composed of disconnected single epochs of EEG recordings to function as a stress test in which technicians must interpret the characteristics of an EEG epoch only from the EEG epoch without temporal relations derived from previous epochs. In these single-epoch test sets, because there are no previous or following epochs to provide information about the current epoch, the technicians can no longer rely on the scoring results from the previous epochs. The intention here was to clearly reveal the effectiveness of the explanations of the EEG characteristics.

In summary, our test data set consisted of two EEG settings: *a set of 15-minute* EEG *segments* and *a set of single-epoch* EEG *segments*. All the participants scored the same set of EEG recordings. A figure explaining our data setting is provided in Multimedia Appendix 1.

Table 1. Summary characteristics of the ISRUC-Sleep Dataset[a] (N=100).

Characteristics	ISRUC-Sleep Dataset
Gender, n (%)	
Male	55 (55)
Female	45 (45)
Age (years), mean (SD)	51 (16)

[a]ISRUC-Sleep Dataset was scored based on American Association of Sleep Medicine Rules.

Experimental Setting

During the experiments, we compared sleep staging performance under 2 different settings. The first was sleep scoring using our CDSS against the baseline AI, where technicians scored stages with AI systems that included only AI predictions provided without any explanation. The second was sleep scoring using our CDSS versus a conventional setting, where technicians need to score each epoch without the predictions by AI. We configured the baseline AI and conventional settings to compare sleep staging settings for our CDSS.

To compare the sleep staging performance under different scoring settings, the technicians had to score each EEG epoch twice as follows: once each with our CDSS and the comparison setting. This was a fair comparison setting to evaluate the efficacy of the system because the characteristics of EEG segments affect sleep staging results significantly. Previous CDSS studies also employed this scoring setting to compare 2 different sleep staging support systems [15]. We divided the test data set into 2 groups and used the first to compare our CDSS with the baseline AI system. A different portion of the test data set was used to compare our CDSS with the conventional sleep staging setting. We randomly permuted the order of the EEG segments and the staging settings. Furthermore, there was a washout period before the second reading of the EEG to avoid the memorization effect.

Quantitative Evaluation

On the basis of the scoring results obtained from the experiments, we evaluated 2 important performance aspects for assessing sleep staging results. First, we considered the accuracy with which the technicians scored the sleep stages under different sleep staging settings. Studies on previous CDSSs have witnessed enhancements in diagnostic accuracy when using the developed CDSS [34-37]. Similarly, we investigated how explanations from our system affect the accuracy of sleep staging. To estimate the classification performance after reviewing the AI predictions, the *macro-F1 score*, which was adopted in previous studies for evaluating sleep staging performance, was used as a performance metric [4,6]. We calculated the metric using the sleep stage labels provided in the public data set as the ground-truth sleep stages. The macro-F1 scores were calculated for each 15-minute EEG segment and a portion of single-epoch EEG recordings.

Second, we evaluated whether interrater reliability was improved by adopting our CDSS. Interrater reliability between polysomnography technicians has been a critical issue in sleep staging because of the variability in interpreting polysomnography recordings among technicians [38]. Following previous work in sleep medicine, which demonstrated that an adequate information system could reduce interrater reliability [19], we investigated whether the information from our CDSS could enhance this property. With this objective, interrater reliability was measured using the *Cohen κ score* [39]. Given the sleep staging results for a 15-minute EEG segment, we calculated the Cohen κ score for every possible pairing of technicians under the same sleep staging setting.

In addition to the above metrics, we also evaluated whether participants could critically assess the accuracy of the model prediction in our system. We calculated the *correction rates of the predictions for incorrectly classified epochs*. Here, we measured the number of incorrectly predicted epochs revised by technicians and incorrectly predicted epochs revised to correct stages. We assumed that for incorrectly predicted epochs, the AI might generate erroneous explanations. Thus, it would be easier for participants to detect incorrectly predicted samples. To evaluate this aspect, we intentionally provided EEG epochs with incorrect AI predictions during the evaluation study.

Qualitative Evaluation

To investigate the extent to which the developed system supported polysomnographic technicians, we conducted semistructured postevaluation interviews. During the survey, we asked questions on a wide range of topics, such as the helpfulness of the information and how the participants adapted to the system. User trust in a system is an important aspect in designing AI-based CDSSs [16,40]. Thus, questions regarding user trust in the developed system were included in the postevaluation interviews. Questions regarding *how information from the system was used in the sleep staging process* were asked during the interviews to reveal notable adoption strategies. The sample interview questions are presented in Textbox 2.

Textbox 2. Examples of interview questions in the qualitative evaluation.

Topic and question statement

User experience of the tool

1. Were the automated predictions and explanations provided in the clinical decision support systems helpful during the experiment? If not, which aspects were unhelpful?
2. How did you perceive the provided explanations when the automated predictions agreed or disagreed with your decisions? Did it affect your trust in the system?
3. Did the explanations correspond well to your perception of the important waveform patterns?

Adoption strategy for the tool

1. How did you use each explanation strategy during the experiment?
2. Was there any notable strategy for adopting the explanations rather than merely accepting the information in the explanations?

Statistical Analysis

As mentioned in the previous section, each EEG epoch was read twice under 2 different settings as follows: once with our CDSS and once with comparison methods, the AI system without explanations, or the conventional staging setting without AI predictions. A statistical comparison was conducted to investigate whether the sleep staging performance was enhanced by adopting our CDSS compared with the comparison settings. Rather than comparing the distribution of the scores, we performed a paired comparison analysis in which we compared 2 sleep scoring performances on the same EEG segments under 2 different score settings. As scoring results could be affected by the complexities and characteristics of particular EEG epochs, it is critical to control these variabilities when assessing the significance of each performance. Furthermore, to exclude variability arising from interrater differences and only consider enhancements in performance by adopting our CDSS, we exclusively performed within-subject analysis for the macro-F1 scores.

The Wilcoxon signed-rank test, a nonparametric statistical test for a set of matched samples [41], was used to estimate the significance of the improvements by adopting the proposed test. For every participant, the data pairs were configured as follows: the macro-F1 and Cohen κ scores from the baseline or usual sleep staging setting (μ_1 κ_1) and the classification results when adopting our CDSS (μ_2 κ_2). For macro-F1 scores, for performance pairs from the same technician, there could be a clustering effect. Thus, we used the Wilcoxon signed-rank test for clustered data, which can account for clustering effects [42-44]. This test aimed to reveal whether performance was significantly enhanced by pairwise comparison when controlling for the variance arising from the interrater characteristics and the differences in EEG epochs. The significance of the results is reported by in terms of *P* values. We set the significance threshold at .05. All statistical and significance tests were performed using Python 3.6. We calculated the *P* values, sample sizes (n), *z* statistics, and effect sizes (*r*) using the Wilcoxon signed-rank test [45].

Ethical Considerations

The study was approved by the institutional review board of the Uijeongbu St Mary's Hospital (IRB number UC20ZADI0137), which waived the requirement for informed consent owing to the nature of the study. All EEG recordings used in this study were acquired from public data sets. All data were anonymized to ensure confidentiality.

Results

Participants Characteristics

In total, 10 polysomnographic technicians were recruited from 3 different affiliations, 2 tertiary hospitals, and 1 secondary hospital. A total of 10% (1/10) of the technicians participated in the user interview, user observation sessions, and an iterative design process. We refer to this participant as technician A throughout the *Results* section. The other 90% (9/10) of the technicians participated in user interviews and evaluation studies. Among the 10 participants, 40% (4/10) were novice technicians with <5 years of experience. A total of 60% (6/10) were senior technicians with >5 years of experience. Technician A, who participated in the tool design process, was excluded from the evaluation study to avoid bias in favor of our CDSS. The participant characteristics are summarized in Table 2.

Table 2. Participant characteristics.

Demographics	Novice technicians (n=4)	Senior technicians (n=6)
Experience (years), mean (SD)	1.75 (1.3)	12.5 (4.7)
Affiliations, n (%)		
Secondary hospital	2 (50)	1 (17)
Tertiary hospital	2 (50)	4 (83)

User Interview: Why Explanation Is Desired

Reasons Technicians Did Not Use Automated Scoring Tools

In total, 20% (2/10) of the participants had no experience of using automatic sleep scoring programs; the other participants preferred not to refer to the automated sleep staging results during sleep staging. The technicians answered that even when predictions were automatically recommended by the software, they removed the automated predictions and scored all the epochs themselves.

In addition to the inaccuracies of algorithms, 50% (5/10) of the participants pointed out that *a lack of explanation* was the main barrier to adopting AI. One technician stated that, "The tools I have experienced do not provide any explanations for predictions, and I need to score every epoch all by myself again when reviewing the predictions." Participants further called for the *clinical soundness of their explanations*. Another technician answered as follows:

> *There certainly exist clinical features to focus on for sleep staging. Even if automatic programs provided some sort of explanation, we need to check whether clinically appropriate EEG features, such as sleep spindles or amplitudes of alpha waves, are used in the algorithms.*

These assertions reflect important considerations regarding explanations and the clinical soundness of algorithm procedures when designing a CDSS [13,16,46].

The Context in Which Explanations Will Be Used

As stated in the subsection above, technicians requested that AI programs should provide clinically sound explanations for predictions, as reviewing the correctness of AI predictions without this information is no different from the manual annotation of sleep stages from scratch. Participants were requested to suggest desirable AI adoption scenarios during the interviews.

In total, 80% (8/10) of the technicians wanted *clinically sound explanations of the predictions*. This is relevant to correct EEG patterns that are important for scoring sleep stages, where users can easily assess the correctness of the reasoning from the AI model based on the conventional manuals for the clinical task:

> *Some automatic programs seem to use procedures that differ from the widely adopted conventions shared among sleep technicians. I think information from AI should adhere to the procedures that we were trained with to make it easier for us to assess the rationale for the explanations.*

Another technician said as follows:

> *When reviewing the AI predictions, I need grounds that convince me. As we are trained to stage based on standard manuals, explanations from AI should be closely related to these processes.*

This point is especially critical in the clinical domain, where predefined sets of rules exist [10].

To summarize the trend of the interview answers, the technicians wanted explanations to validate the correctness of the AI predictions based on their clinical knowledge of sleep staging.

User Observation: What Information Should Be Contained in Explanations

By observing technician A for 1 month, we obtained an understanding of how technicians interpret EEG signals during sleep staging. Using the clinical context proposed in the manual [1], we categorized EEG patterns based on how the technician processed the information in the EEG recordings. On the basis of how they processed each EEG feature, we created a list of explanation types that can be provided in the CDSS. The candidate explanation-type categories are listed in Textbox 3.

Textbox 3. Explanation type to be provided in the clinical decision support systems.

Explanation type 1: occurrence of signals

For some patterns in electroencephalogram recordings, their presence is a clear indicator of certain sleep stages. For example, the occurrence of *sleep spindles* and *k-complexes* is strongly correlated to non–rapid eye movement (REM) 2 stages. In general, technicians search the entire signal to find these patterns. Therefore, proper detection of these patterns is sufficient information for polysomnographic technicians.

Explanation type 2: ratio of signals

Technician A claimed that estimating the ratio of *delta waves* in an epoch is the most critical part in identifying the non-REM 3 stages since the scoring manuals recommend annotating the epoch as stage non-REM 3 when delta waves account for more than 20% of the signals [1]. The participant mentioned that technicians usually count the number of delta waves manually to correctly identify the non-REM 3 stages in sleep recordings.

Explanation type 3: changes in signals

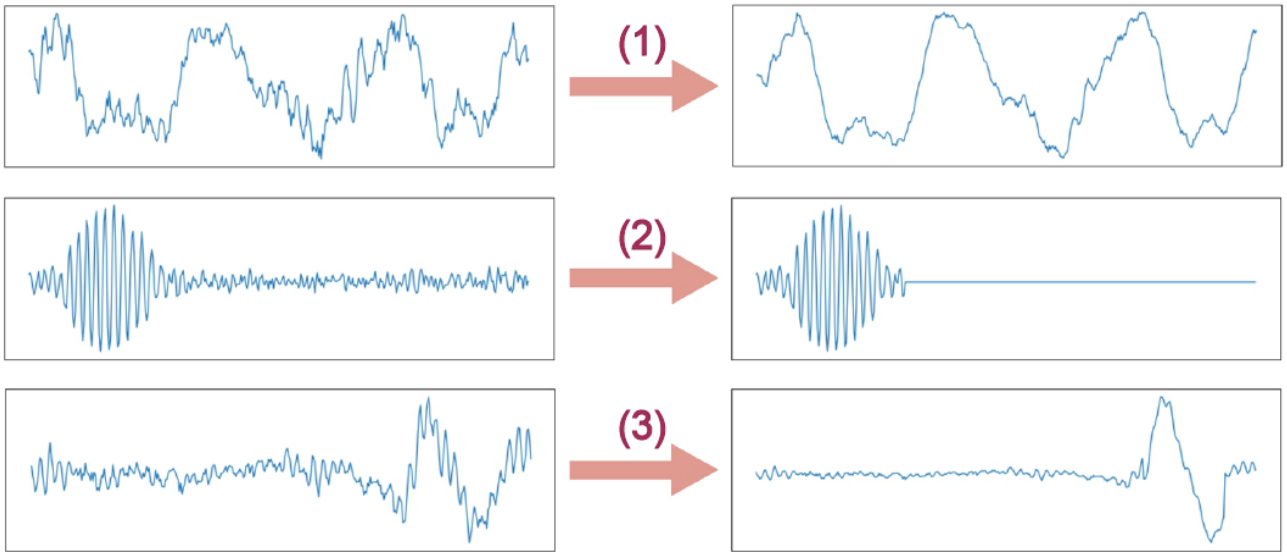
Alpha waves are prevalently observed during the wake and non-REM 1 stages. However, the participant mentioned that changes in the amplitudes of *alpha waves* are important criteria for distinguishing non-REM 1 stages from the wake stages. According to the manual [1], the alpha waves in the non-REM 1 stages normally exhibit smaller amplitudes compared with the wake stages. Technician A mentioned that perceiving the overall changes in alpha waves is the primary task in detecting boundaries between the wake and non-REM 1 stages.

Iterative Design Process: How Explanations Can Be Presented

Refinements of Model Components

In the first iteration session, convolutional filters obtained from TRIER [28] were shown to technician A. The participant expressed the concern that although the convolutional filters contained morphologically significant shapes, undesirable features (high-frequency noises or low-frequency fluctuations) were also intermingled in the filter. The participant requested a refinement of the convolutional filters to improve the quality of the features. For example, in formulating filters that correspond to slow waves, the participant wanted to remove high-frequency components because delta waves have frequency components <4 Hz. The filter refinement process is illustrated in Figure 2. Consequently, the convolutional filters contain features that correspond to the following EEG patterns: alpha waves, theta waves, delta waves, sawtooth waves, vertex sharp waves, sleep spindles, and k-complexes. After refinement, the filters are depicted in Multimedia Appendix 1.

Figure 2. The filter refinement process is as follows: (1) delta waves were low-pass filtered, (2) regions outside the sleep spindle were zeroed-out, and (3) only the regions corresponding to k-complex features were selected and low-pass filtered afterward.



Selecting Information Source for Making Explanations

Owing to the previous refinement process, components in the convolutional filter are clinically meaningful, and the corresponding features in the neural networks can be interpreted accordingly. For example, for a filter that was designated for k-complex–related features, the activation values generated from the filter were used to locate k-complexes in the data. Similarly, filters analogous to alpha waves can generate information related to alpha wave changes in the data.

Therefore, we selected convolutional filters and activation values as basic elements to generate explanations of the model predictions. In addition to the 2 components, a saliency map [29], or the gradient values of the input points, was also adopted to mark significant regions in making a prediction. This information indicates which regions in the data were important from the AI perspective. The neural network components used for generating explanations are summarized in Textbox 4.

Textbox 4. List of information sources for generating information for the clinical decision support systems.

Component 1: convolutional filters and their activation values

Convolutional filters represent the clinical electroencephalogram patterns on which the model is based. Information regarding each clinical feature can be obtained from the activation values acquired from the filters.

Component 2: saliency values calculated from neural networks

Important regions, which significantly contributed to model predictions, can be inspected from the saliency values. Users can view the data from the perspective of the artificial intelligence model with saliency values.

Visualization Strategies

Visualization strategies for each clinical feature were devised to provide information in an easily adopted form for sleep staging. Initially, plots of activation vectors without any processing were provided to the participating technician. In this case, the technician failed to use any of the information in the activation values. They emphasized that information should be compatible with the scoring procedure of the technician: “I cannot make use of the information. I want information to be provided in a form that can easily fit with my procedure.” This argument is closely linked to the critical issues in designing AI assistant tools: information from the system should be easily integrated into tasks of users [47,48].

From this standpoint, we constructed different visualization strategies for each explanation type because conventions observed during the user observation study constituted the

representative logical procedures for processing information in EEG recordings (Textbox 5).

Figure 3 shows an in-tool visualization of the strategies. Through visualization, explanations from AI can be conveyed to users with their proper clinical contexts. Technician A attested that such explanations with enhanced readability could be easily adopted in the sleep staging process.

Textbox 5. Four visualization strategies developed in this study. The first three strategies correspond to the interpreting conventions observed during the user observation study.

Strategy 1: detection boxes

Technician A claimed that the patterns, the presence of which alone indicates a sleep stage, should be more easily identified from the recordings. After that, it would be sufficient for the technicians to check whether the artificial intelligence (AI) model correctly located these patterns in the electroencephalogram (EEG) recordings. Therefore, we outlined detection boxes in regions that were detected to include the desired EEG patterns. Detection algorithms were implemented based on the amplitudes of the activation values calculated from the convolutional filters with the desired pattern.

Strategy 2: delta wave blocks

As polysomnographic technicians rely on the number of delta waves in the recordings, it is important to make the distribution of delta waves more visually intuitive. For these cases, technician A wanted to perceive each peak in delta waves as a single entity. We digitized the activation values from the convolutional filters of delta waves such that regions with activation values higher than a set threshold were encoded as 1 or otherwise as 0. Visualizing the encoded digits from the activation vectors, technician A perceived the information as blocks of slow waves and counted the number of blocks in the figure.

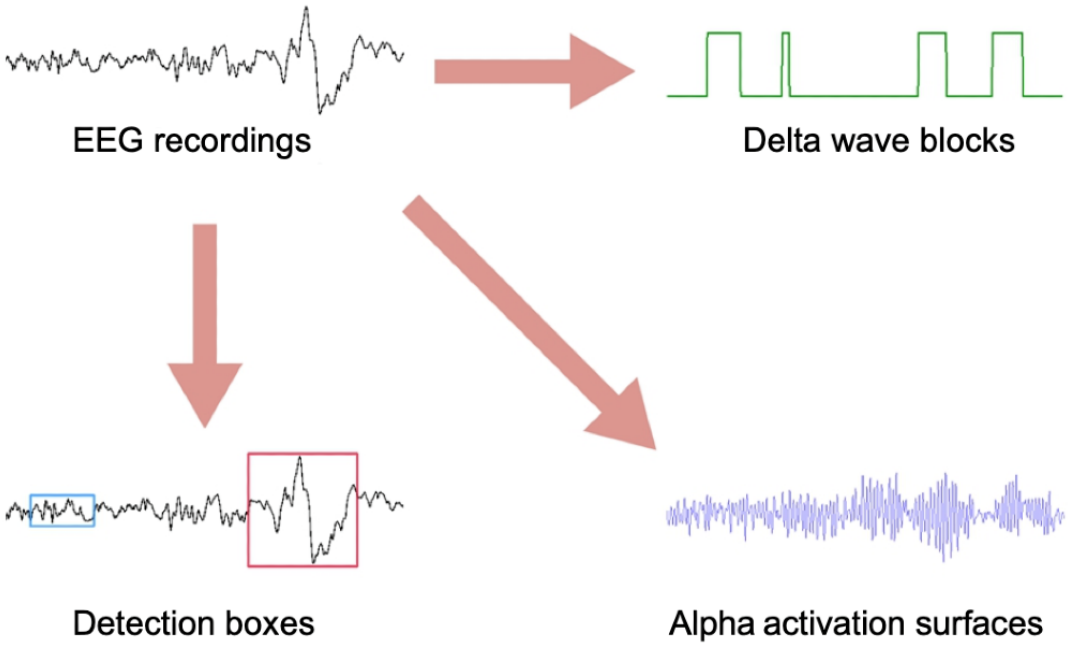
Strategy 3: alpha activation surfaces

For detecting changes of alpha waves on the boundary of the wake and non-REM 1 stages, the activation values generated directly from the convolution between the alpha wave filters and the input recordings were used. In this case, the participant requested fluctuations to be easily perceivable in the interface. During the iterations, technician A acknowledged that overall fluctuations of the activation values matched well with the perception of the changes. In such a setting, it was felt that the activation values amplify the changes in amplitudes. The technician asserted that these values are perceived as a surface area, thus making it more intuitive to sense overall changes in the signal.

Strategy 4: saliency highlights

The participant claimed that saliency values could be helpful for technicians as they could view the recordings from the AI perspective. In particular, the technician wanted to identify the EEG regions with high saliency values. Therefore, we highlighted the EEG recording segments with high saliency values.

Figure 3. Visualization strategies for each interpretation pattern. Information in electroencephalogram (EEG) recordings is visualized differently for each interpretation convention introduced in Textbox 5.



Constructing the System Outline

In the original version of the system, we empower users to explore EEG recordings interactively with a filter selection box with which users could choose the desired EEG patterns and analyze signals based on the selected features. However, technician A observed that the system with an exploratory filter selection process might degrade its usability, as it disrupts the workflows of the technician:

> *Usually scoring of an epoch takes place in a short time, typically between five to ten seconds and even down to one second for easy cases. The selection process can be a bottleneck during scoring, thus other technicians are more likely to skip filter selection and score EEG epochs on their own.*

This indicates that for clinical tasks where large numbers of data points are annotated in a relatively short time, the accessibility of desired features could be more important than interactivity. Therefore, instead of interacting with multiple features, we implemented an information system to be directly accessible.

Specifically, rather than providing multiple sets of available information, we chose to show only the information corresponding to the predicted sleep stage for the epoch (Figure 4). For example, only the detection boxes of sleep spindles and k-complexes were provided for the epochs that were predicted as N2 stages. In this version, technician A acknowledged that the usability of information is enhanced compared with previous versions where multiple sets of information are provided, which results in too much information on a single screen and poor readability. Furthermore, the visualizations could explain the model predictions because the model provides only information relevant to its predictions. In Table 3, we list specific information provided for each stage.

Similar to other tools for assisting sleep staging [49], our system provides basic information from EEG recordings (Figure 5). It displays the hypnogram, a graph that visualizes changes in sleep stages over time, on top of its interface. Hypnograms for annotated stages from users as well as predictions from AI are provided so that users can monitor their editing process. A table that contains time information and annotated sleep stages is located on the right panel of the interface. The EEG and EOG recordings of an epoch are depicted in the main interface. In addition to the basic components, our CDSS provides the following information: AI-generated predictions and explanations from the AI model around the target EEG channel. Video recording provided in Multimedia Appendix 2 demonstrates the overview of the CDSS and how users interact with it.

Figure 4. Visualization strategies for the system. In the electroencephalogram (EEG), the recordings predicted as N2, k-complex, and sleep spindles are detected and visualized as red and blue boxes. In EEG recordings predicted as N3, detected delta waves are visualized as green blocks. Regions with high saliency values are highlighted in pink on the EEG recordings. Strategy is abbreviated as S.

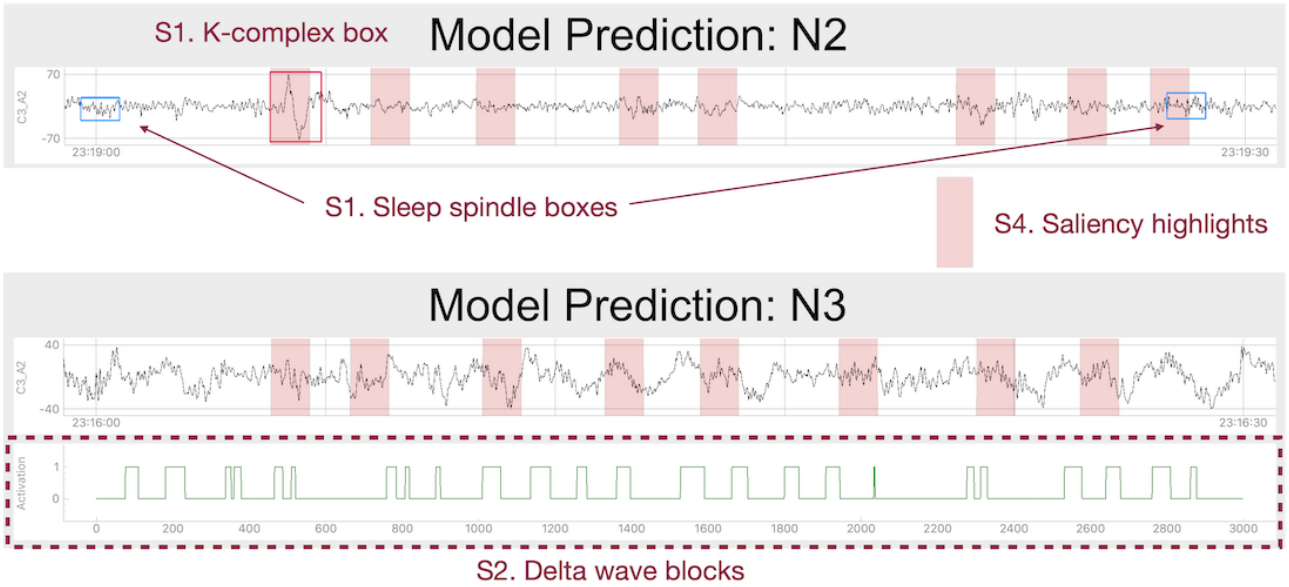


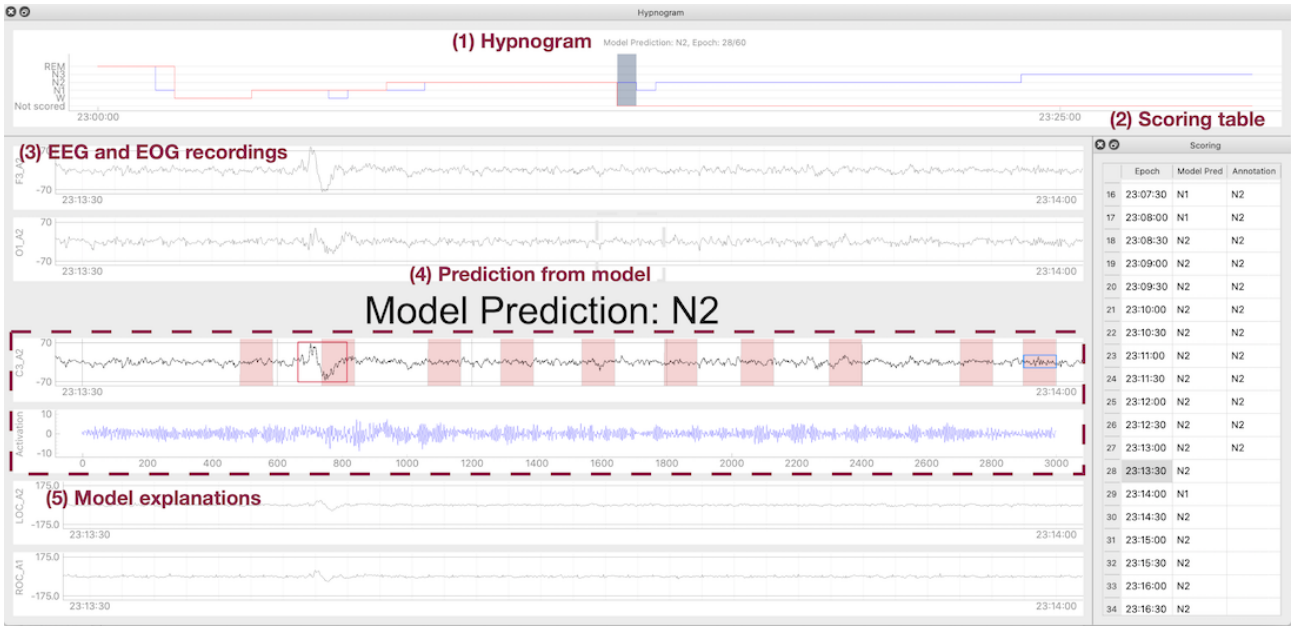
Table 3. Information provided for each sleep stage.

Stage	Detection boxes	Delta wave blocks	Alpha activation surfaces	Saliency highlights
Wake			✓	✓
N1[a]			✓	✓
N2[a]	✓		✓	✓
N3[a]		✓		✓
REM[b]	✓		✓	✓

[a]N1-3: non–rapid eye movement stages 1-3.

[b]REM: rapid eye movement.

Figure 5. The following is the overall interface of the system: (1) hypnogram; (2) scoring table lists the time sequence of model predictions and user annotations; (3) physiological recordings of the data set are visualized in the main panel; (4) predictions; and (5) explanations from artificial intelligence (AI) are in the middle of the interface. EEG: electroencephalogram; EOG: electrooculogram.



Quantitative Evaluation

Accuracy

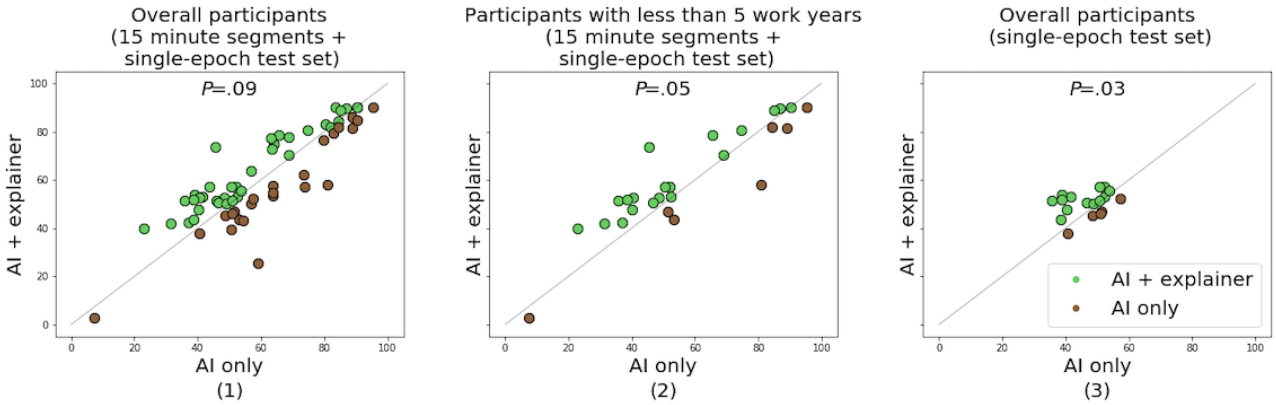
Figure 6 illustrates macro-F1 scores. Each point in the scatter plots corresponds to the performance pair measured using the comparison method (AI only, μ_1) and our method (AI+explainer, μ_2) on the same test set.

For the overall data set, which consisted of 15-minute EEG segments and single-epoch test set, there were no significant differences between baseline AI and our CDSS for results from all participants (μ_1=60.22; μ_2=61.31; P=.09; n=26; z=1.63; number of clusters=9). However, a performance improvement can be observed when we restricted this data set to participants with <5 years of work experience (μ_1=56.75; μ_2=60.59; P=.05; n=26; z=1.63; number of clusters=4). For a single-epoch test set, in which the utility of the methods could be more accurately determined, we also observed improvements in accuracy (μ_1=46.55; μ_2=50.28; P=.03; n=18; z=1.94; number of clusters=9).

For the overall data set, compared with the conventional staging setting where predictions from the AI were not provided (μ_1), the macro-F1 scores were significantly improved when the technicians adopted our method (μ_1=43.23; μ_2=68.04; P=.004; n=17; z=2.64; number of clusters=9). Similarly, the macro-F1 scores improved for novice technicians when we compared our CDSS with a conventional sleep staging setting (μ_1=39.52; μ_2=70.58; P=.05; n=6; z=1.67; number of clusters=3).

It should be noted that these results cannot be directly compared with sleep staging performance in other studies where performance was evaluated for whole-night sleep staging results. In our setting, the performance was measured from short segments of the EEG recordings. Here, sleep staging performances could be reported to be lower than the whole night sleep staging results in previous works, as the macro-F1 scores of the sleep staging results could be significantly affected by a few incorrect predictions.

Figure 6. The improvements of the macro-F1 scores in various settings. The results measured as follows from (1) all participants and all test sets; (2) participants who have <5 years of work experience and all test sets; and (3) all participants and single-epoch test sets are provided. AI: artificial intelligence.



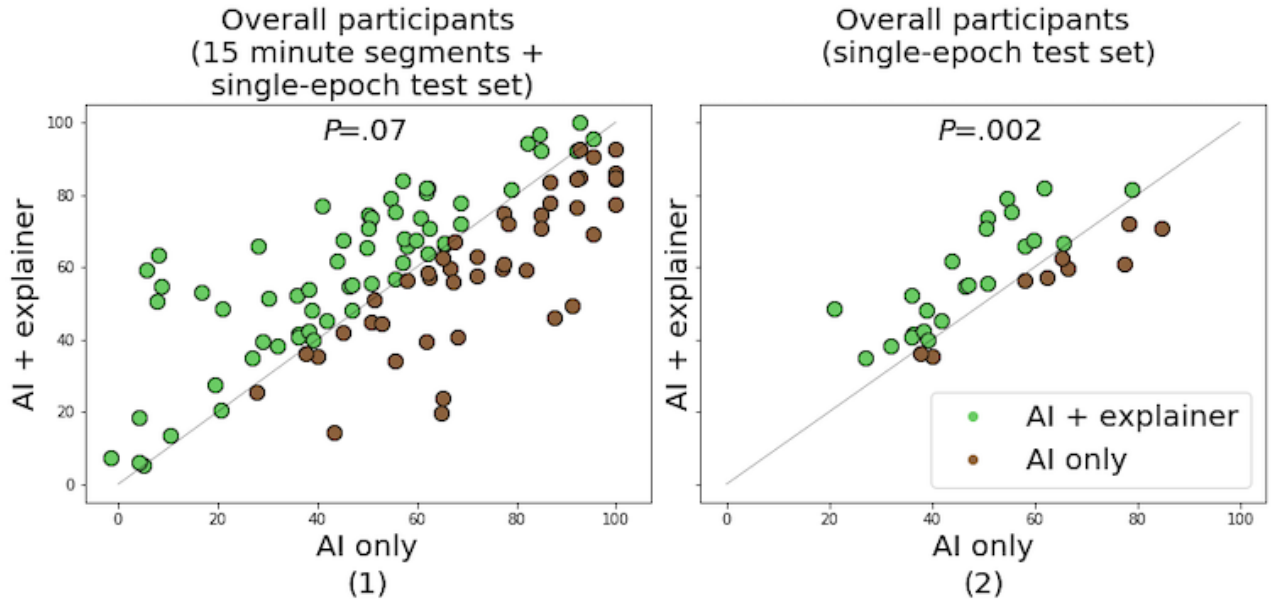
Correction Rates for Incorrect Predictions From the AI

For the erroneous predictions generated by the AI, statistics regarding the ratio of correctly revised epochs did not show significant differences between the baseline AI and our method. Among the 392 EEG epochs in the test data, 30.8% (121/392) were incorrectly predicted epochs from our network. Of the 30.8% (121/392) of the epochs, technicians detected 28.5% (112/392) of the incorrect predictions made with our CDSS, whereas 28.5% (112/392) were detected in the baseline AI. There were no significant differences in the detection rates of incorrectly predicted epochs (P=.39; n=9; z=0.28;r=0.11). Furthermore, among these incorrectly predicted epochs from AI detected by technicians, there were no significant differences in the ratio of correct revisions where technicians identified the correct stages for incorrect predictions (μ_1=15.68%; μ_2=16.42%; P=.76; n=9; z=−0.70; r=0.28). Similarly, for technicians with <5 years of experience, we did not observe improvements in the detection rates of incorrectly predicted epochs (μ_1=27.19%; μ_2=30.52%; P=.86; n=9; z=−1.10; r=−0.60) and the ratio of correct revisions (μ_1=12.90%; μ_2=16.67%; P=.97; n=9; z=−1.83; r=−1.0).

Interrater Reliability

Scatter plots of the Cohen κ scores calculated for the baseline (κ_1) and our method (κ_2) are shown in Figure 7. As with the macro-F1 scores, improvements in reliability for all cases were observed (κ_1=57.02; κ_2=59.54; P=.07; n=212; z=1.49; r=0.17). However, more significant improvements were observed for the single-epoch test set (κ_1=51.28; κ_2=57.21; P=.002; n=64; z=2.80; r=0.57). According to the criteria for interpreting the Cohen κ score [50], we obtained moderate agreement between technicians for both the proposed CDSS and baseline AI settings. Compared with usual sleep staging settings, where predictions from AI are not provided, interrater reliability also improved (κ_1=35.06; κ_2=77.48; P<.001).

Figure 7. The improvements of interrater reliability in various settings. The results measured from the following: (1) all participants and all test sets and (2) all participants and single-epoch test sets are provided. AI: artificial intelligence.



Qualitative Evaluation

In this section, a qualitative evaluation of the tool is described. The adoption strategies developed by the participants and the perceived usability of the system are discussed.

Helpful Aspects

In total, 78% (7/9) of the participants responded that our system helped to review AI predictions. They reported that they referred to information from the CDSS when inspecting AI predictions. Several aspects of the utility of the tools were confirmed.

Reducing the Workload Required for Pattern Recognition

One of the most important utilities mentioned during the interviews was that our tool reduced the workload required to inspect EEG epochs. Analyzing EEG epochs is similar to visual searching tasks, where technicians must identify specific patterns in a visual environment [51]. Participants attested that information visualized by saliency highlights and detection boxes drew their attention to important regions [52]. Helped by the information provided by the detection boxes, participants were easily able to identify important regions for examination. On the basis of this information, they assessed whether the patterns were correctly detected by the algorithm. Similarly, for delta waves, participants replied that they only needed to count the number of delta wave blocks, and they did not need to check delta waves one by one from the EEG recordings.

Providing Quantitative Visual Reference

Interviewees stated that sleep staging tasks heavily depend on the subjective criteria of each technician. Perceiving the attenuation in alpha waves on the boundary of the wake and N1 stages is one of the most representative cases in which sleep staging is affected by subjective perception. In total, 55% (5/9) of the participants used the information from the alpha activation surfaces as a reference when they were not confident whether changes in the alpha wave were significant. Even the participant who was not satisfied with the system answered that this information was helpful for similar reasons.

Unhelpful Aspects

Two senior participants answered that they did not find the system helpful. They claimed that the specificity of the information from the system was below the desired level:

> *I am quite strict in detecting sleep EEG patterns like delta waves. However, from my point of view, too many regions were annotated as significant points. Thus, for many cases, I did not refer to the provided information.*

This point emphasizes that for clinical tasks where decision-making may differ between individuals, personal differences among users should be considered to improve the usability of a tool. In our domain, for example, user interfaces that control the sensitivity and specificity of the pattern detection algorithm can be provided.

Another technician did not refer to the system during the experiments because it was inconvenient to consider information other than the EEG recordings:

> *Due to time constraints, I am used to scoring stages speedily compared to other technicians. Thus, in some sense, I tend to rush during sleep staging sessions, and would rather not care about information in the system.*

Interviews from the participants reveal that the tight time constraints in clinical environments are another challenge to be considered when designing a clinical support tool because changing the workflow of medical staff is a complicated task, which requires not only reliable performance but also usability in the workflows [46].

Explanations and Trust in the System

In this section, we summarize how the explanations of our systems affect user trust during the experiments.

Explanations in Agreed Epochs

For epochs in which the predictions of the participants agreed with those of the AI, the technicians expressed trust in the predictions. In this case, the participants expressed that, as annotated regions from the system matched the important regions determined by the users, they were confidently able to continue to the next stages.

Explanations in Epochs With Disagreement

For the epochs where the predictions differed between the AI and users and were consequently modified by the technicians, the participants felt that the explanations clarified why AI predicted the epochs differently. In these cases, one technician argued:

> *Without explanations, I might jump to a conclusion that the accuracy of the AI is not at a desirable level. However, after being exposed to the explanations, there were some convincing factors in the AI-generated predictions, and I tried to re-investigate the recordings based on the AI explanations to find out whether my reasoning on predictions was strong enough to modify the AI prediction.*

Even the technician, who did not think that the tool was helpful, reported:

> *At first, I totally disagreed with the predictions from the AI. Throughout the experiments, however, I found out that AI algorithms were reasonable on some level.*

In summary, even though user trust could be severely affected when the AI predictions were inconsistent with those of the users, the explanations provided in our CDSS improved the trustworthiness of the system. In particular, the explanations helped users find reasonable aspects of AI predictions.

Notable Adoption Strategies

We obtained various sets of answers such as “I first focused on saliency highlights and then inspected signals based on the detection boxes” or “I used the alpha activation surfaces in detecting sleep arousal.” Among these answers, some notable strategies were identified. We discuss these strategies and their implications for human–AI collaboration in clinical domains.

Rediscovery of Unnoticed Features

Classifying REM stages solely from EEG recordings is deemed an impossible task, and sleep technicians prefer to rely on chin EMG and EOG recordings for REM stages [1]. Thus, most participants had difficulty evaluating sleep epochs that were predicted as REM. However, several participants found that they could distinguish REM from the N1 stages with our method:

> *In general, I used to disregard sawtooth waves because REM has more distinct landmarks in EOG. However, the AI model correctly captured the sawtooth waves (patterns that occur in REM stages) and convinced me that the given epochs are from REM. Without such information, I might incorrectly score the stages.*

These use cases demonstrate that our tool successfully conveyed important but easily dismissed features of the data. We believe that the above insight illustrates an important aspect of human–AI collaboration because alternative but significant viewpoints from the AI system successfully convinced the users during decision-making, which resulted in a performance enhancement.

Attention Allocation

In the adoption of a clinical AI system, to allocate their attention efficiently to weak portions of the algorithms, it is important for users to properly understand the strengths and weaknesses of AI. This scenario is termed the attention allocation [18]. During the experiments, several technicians developed strategies related to attention allocation. One participant found that AI is vulnerable to misidentifying sweat artifacts as delta waves. This participant strategically allocated more attention to annotated regions in epochs that were predicted as N3 stages and inspected whether the annotated regions corresponded to delta waves or sweat artifacts. With this strategy, this participant effectively distinguished the N3 stages from epochs contaminated by sweat artifacts.

In this adoption pattern, participants constructed strategies to successfully collaborate with AI [48]. Specifically, users evaluated the convincing and unconvincing contributions of AI, thus efficiently allocating their attention during the adoption.

Discussion

Principal Findings

To our knowledge, this work is the first to construct an interpretable AI system using deep learning with a user-centered approach to develop a CDSS for sleep staging. Recent studies continuously demonstrate that deep learning algorithms can achieve comparable performance compared with human experts [7-9]. However, previous studies have found that human practitioners require information beyond the delivery of accurate predictions [18]. To achieve this, we focused on constructing a CDSS that provides information compatible with the diagnostic patterns of human raters and helps technicians easily integrate the CDSS into their sleep staging procedures. Through user observation and an iterative design process, we obtained the desired characteristics for the explanations provided in the CDSS for sleep staging. First, clinical practitioners wanted explanations to help them validate AI predictions. Here, technicians wanted explanations that adhered to their clinical knowledge. Second, we categorized the type of information based on our observations of how technicians interpret the characteristics of each EEG. Finally, during the iterative design process, we confirmed that information contained in neural network components can be used to generate explanations for sleep staging results. The design components were updated iteratively based on the feedback of the technician.

When evaluating the improvements in the sleep staging performance of all participants, we did not observe significant improvements when the *P* value was approximately .17. However, we believe that our quantitative evaluation contains meaningful results. First, when assessing the improvements for novice participants, we observed that the macro-F1 scores improved by 6.7% with a *P* value of .02. Considering that novice technicians may rely more on supportive information than expert technicians, this result implies that our tool could be effectively used to augment the sleep scoring capacities of novice technicians with acceptable sleep-relevant explanations. Second, when assessing the improvements in a single-epoch sleep scoring setting, which is similar to a stress-test configuration, we observed significant improvements in the macro-F1 scores and interrater reliability. Notable results in this stress test setting could indicate that our explanations to an extent helped technicians interpret the signal characteristics of each EEG epoch. Third, the results of the qualitative evaluation implied that the CDSS supports sleep staging by reducing the workload required for pattern recognition and providing quantitative visual references. These findings show that the developed system successfully and appropriately complemented the assessments of the technician by suggesting the desired information. Our tool obtained such utility for two reasons: (1) clinically sound features were correctly addressed and (2) information visualization was designed to be acceptable in conventional workflows of the sleep staging process.

We identified further issues that should be considered when designing a CDSS. During the experiments, 20% (2/10) of the technicians indicated that our system was not adoptable for workflow in sleep staging. In particular, 10% (1/10) of the technicians expressed a lack of trust in the AI system. In general, the avoidance of algorithmic results is an important challenge to be addressed when adopting an automatic system [53]. However, these challenges can be interpreted based on skill levels of the technician. For example, based on the Rasmussen skill-, rule-, and knowledge-based behavior model [24], senior technicians may score sleep stages without consciously processing EEG information. Therefore, additional explanations from the CDSS can distract such technicians. In contrast, novice technicians may require additional cognitive processing of information in the recordings. Therefore, explanations from the CDSS could be helpful as guiding information during processing and lead to significant enhancements in their performance when a CDSS is adopted.

In addition, over reliance on computer systems is another challenge to be considered when adopting decision support tools

[11,54]. When adopting AI systems, there are cases where users tend to accept predictions from systems without any personal judgment on whether the information is correct. In our evaluation, the correction rates for erroneous predictions did not improve. This means that even though explanations from our system successfully operated as convincing components for model predictions, they failed to reveal ambiguous predictions. These results have implications for further development (eg, explanations for uncertainties in predictions can be provided by the model to inform users about ambiguous components in the data [15]). The confidence of the predictions can be algorithmically estimated by the models as additional information [55]. Such features can be integrated into a single framework to enhance safety in human–AI interaction systems.

Comparison With Previous Work

Previous studies on sleep staging have confirmed that suggestions for proper computational features can enhance sleep staging performance. An experimental study demonstrated that interrater reliability among technicians can be significantly improved by computer-derived suggestions [19]. Taking inspiration from that study, our work proposes an approach to provide clinically meaningful information from deep learning models. Our results are consistent with those of a previous study, as the interrater reliability in our system improved significantly. However, our study differs from previous works in several respects. Although previous tools for sleep staging have already provided sleep-relevant information to users [56,57], these algorithms require a large amount of parameter tuning to fit each data set [58]. In this sense, these works used a manually curated algorithm rather than augmenting the AI system to provide information. Furthermore, our work addresses the utility and readability of the system during the development of the tool, whereas previous studies preferentially focused on the calculation of sleep-relevant features in EEGs.

In the domain of human–AI interaction, several deep learning models have been exploited as information sources to assist medical staff with appropriate knowledge. In these works, the usability of clinical AI was mainly addressed from the perspective of human users [18]. A previous study surveyed how and what information should be provided for the analysis of radiographic images [59]. This work stressed that information systems should be designed based on the user needs of clinical practitioners. Another study introduced a novel medical image retrieval system that leverages embedding vectors in a neural network to retrieve similar medical images [47]. These bodies of work demonstrated that model interpretations should be formulated in the context of clinical knowledge, as users require medical explanations during adoption. Similarly, our work extensively investigates the desirable characteristics of sleep staging AI and proposes how these features can be provided in a CDSS.

For sleep staging, an earlier work proposed an AI framework that prioritizes ambiguous epochs in EEG recordings with explanations in cases of uncertainty [15]. However, this study proposes a conceptual framework rather than a practical implementation of the system. In this work, CDSS was simulated in a Wizard of Oz experiment, where human researchers manually generated the explanations in the system to address the ambiguous epochs in EEG recordings. In contrast, our work proposed a practical methodology for constructing meaningful information on sleep stages to assist clinical practitioners.

Limitations and Future Directions

The limitations that require consideration remain in our study. First, we conducted user observations and iterative design sessions with only 1 technician. Although manuals for sleep staging support most of the feedback of the technician, specific requirements defined by different users are necessary for user-centered design research. Moreover, during the experiments, participants reviewed the EEG recordings provided from a public EEG data set. As EEG recordings are highly heterogeneous across data sets and recording environments, the utility of the system could be more accurately evaluated if the neural network model was trained on data sets recorded in real-world settings.

Our work is further limited as we only considered EEG recordings for sleep scoring. Assuming real-world sleep scoring is performed with polysomnographic recordings, which include EEG, EOG, EMG, and ECG signals, not considering other recordings may have affected the scoring results. For example, eye movement patterns are crucial factors in identifying the REM stages. As we have only provided information for EEG recordings, we could not offer explanations regarding eye movements. However, we believe that our overall design approach can be applied similarly in future studies to explain the output of other physiological sensors, such as EOG and EMG. These future studies could construct a more comprehensible CDSS for sleep scoring. In addition, evaluation of the CDSS system with whole-night polysomnography will provide more generalizable performance results that can be connected to the results of real-world polysomnography.

The overall sample size may not be sufficient for comparison, considering that there are high interrater disagreements on the sleep staging results depending on individual characteristics. Even though we observed some notable improvements with the small sample size, a further evaluation study with more technicians is desirable. Furthermore, the representativeness of participants should be mentioned. Technicians from secondary and tertiary hospitals participated in the evaluation study, and technicians in primary hospitals were not considered. Technicians in primary hospitals may exhibit different tendencies toward the adoption of automatic sleep scoring tools. Thus, our study did not address this population. However, considering that technicians in primary hospitals tend to have relatively short experience in polysomnography, we believe that these results from novice technicians can be generalized to polysomnographic technicians in primary care.

An AI system that provides explanations for predictions was compared with conventional models that do not provide explanations. In this setting, there was a risk that the participants were aware that the experimental objective was to construct and evaluate the effectiveness of the explanations. However, considering that explainable AI systems for medical domains have not been widely developed, many previous CDSS studies conducted experiments in a similar manner to our work [15].

Nevertheless, the omission of blinding conditions is a limitation of our experimental setting.

Although our work qualitatively evaluates how users perceive the CDSS, future work is required to quantitatively assess the usability of the tool. For example, the NASA-Task Load Index [60] could be used in a prospective study to compare the required workload for each sleep scoring tool. Other aspects, such as time spent scoring sleep stages, could be estimated in a more controlled experimental setting. We believe that future studies will provide more insights into the usability of CDSS.

Conclusions

Our findings indicate that formulating clinical explanations for automated predictions using information from an AI system that incorporates a user-centered design process is an effective strategy for developing a CDSS for sleep staging. The proposed CDSS has great potential to be integrated into the real-world clinical workflow in a sleep laboratory based on the extent to which performance was improved and is highly useful in sleep staging.

Acknowledgments

This study was supported by Looxid Labs, Korea, and a grant of the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (grant HI21C0852). All the code and data sets used in this study are available on GitHub.

Authors' Contributions

All authors conceived the study, participated in the implementation of the tool, and wrote the manuscript. JH and TL conducted user interviews and user observation studies.

Conflicts of Interest

None declared.

Multimedia Appendix 1
List of sleep-relevant electroencephalogram (EEG) patterns, refined convolutional filters, constructing data sets with EEG segments, and convolutional neural network components.
[DOCX File , 108 KB - jmir_v24i1e28659_app1.docx]

Multimedia Appendix 2
Demo video of the clinical decision support system.
[MP4 File (MP4 Video), 29384 KB - jmir_v24i1e28659_app2.mp4]

References

1. Berry R, Brooks R, Gamaldo C, Harding S, Lloyd R, Marcus C, et al. The AASM manual for the scoring of sleep and associated events: rules, terminology and technical specifications. American Academy of Sleep Medicine, Darien, IL. 2015. URL: http://aasm.org/resources/pdf/scoring-manual-preface.pdf [accessed 2021-12-29]
2. Subramanian S, Hesselbacher S, Mattewal A, Surani S. Gender and age influence the effects of slow-wave sleep on respiration in patients with obstructive sleep apnea. Sleep Breath 2013 Mar 16;17(1):51-56. [doi: 10.1007/s11325-011-0644-4] [Medline: 22252284]
3. Andlauer O, Moore H, Jouhier L, Drake C, Peppard PE, Han F, et al. Nocturnal rapid eye movement sleep latency for identifying patients with narcolepsy/hypocretin deficiency. JAMA Neurol 2013 Jul;70(7):891-902. [doi: 10.1001/jamaneurol.2013.1589] [Medline: 23649748]
4. Supratak A, Dong H, Wu C, Guo Y. DeepSleepNet: a model for automatic sleep stage scoring based on raw single-channel EEG. IEEE Trans Neural Syst Rehabil Eng 2017 Nov;25(11):1998-2008. [doi: 10.1109/TNSRE.2017.2721116] [Medline: 28678710]
5. Perslev M, Jensen M, Darkner S, Jennum P, Igel C. U-Time: a fully convolutional network for time series segmentation applied to sleep staging. arXiv. 2019. URL: https://arxiv.org/pdf/1910.11162.pdf [accessed 2021-12-29]
6. Phan H, Andreotti F, Cooray N, Chen OY, De Vos M. SeqSleepNet: end-to-end hierarchical recurrent neural network for sequence-to-sequence automatic sleep staging. IEEE Trans Neural Syst Rehabil Eng 2019 Mar;27(3):400-410 [FREE Full text] [doi: 10.1109/TNSRE.2019.2896659] [Medline: 30716040]
7. Sokolovsky M, Guerrero F, Paisarnsrisomsuk S, Ruiz C, Alvarez S. Human expert-level automated sleep stage prediction and feature discovery by deep convolutional neural networks. In: Proceedings of the 17th International Workshop on Data Mining in Bionformatics (BIOKDD2018), in Conjunction with the ACM SIGKDD Conference on Knowledge Discovery and Data Mining KDD2018. 2018 Presented at: 17th International Workshop on Data Mining in Bionformatics (BIOKDD2018), in conjunction with the ACM SIGKDD Conference on Knowledge Discovery and Data Mining KDD2018; Aug 20, 2018; London, UK URL: https://tinyurl.com/4ptkavfv

8. Tschandl P, Rosendahl C, Akay BN, Argenziano G, Blum A, Braun RP, et al. Expert-level diagnosis of nonpigmented skin cancer by combined convolutional neural networks. JAMA Dermatol 2019 Jan 01;155(1):58-65 [FREE Full text] [doi: 10.1001/jamadermatol.2018.4378] [Medline: 30484822]
9. Hannun AY, Rajpurkar P, Haghpanahi M, Tison GH, Bourn C, Turakhia MP, et al. Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. Nat Med 2019 Jan;25(1):65-69 [FREE Full text] [doi: 10.1038/s41591-018-0268-3] [Medline: 30617320]
10. Shortliffe EH, Sepúlveda MJ. Clinical decision support in the era of artificial intelligence. J Am Med Assoc 2018 Dec 04;320(21):2199-2200. [doi: 10.1001/jama.2018.17163] [Medline: 30398550]
11. Magrabi F, Ammenwerth E, McNair JB, De Keizer NF, Hyppönen H, Nykänen P, et al. Artificial intelligence in clinical decision support: challenges for evaluating ai and practical implications. Yearb Med Inform 2019 Aug;28(1):128-134 [FREE Full text] [doi: 10.1055/s-0039-1677903] [Medline: 31022752]
12. Smith H. Clinical AI: opacity, accountability, responsibility and liability. AI Soc 2020 Jul 25;36(2):535-545. [doi: 10.1007/s00146-020-01019-6]
13. Goldstein CA, Berry RB, Kent DT, Kristo DA, Seixas AA, Redline S, et al. Artificial intelligence in sleep medicine: background and implications for clinicians. J Clin Sleep Med 2020 Apr 15;16(4):609-618 [FREE Full text] [doi: 10.5664/jcsm.8388] [Medline: 32065113]
14. Goldstein CA, Berry RB, Kent DT, Kristo DA, Seixas AA, Redline S, et al. Artificial intelligence in sleep medicine: an American Academy of Sleep Medicine position statement. J Clin Sleep Med 2020 Apr 15;16(4):605-607 [FREE Full text] [doi: 10.5664/jcsm.8288] [Medline: 32022674]
15. Schaekermann M, Beaton G, Sanoubari E, Lim A, Larson K, Law E. Ambiguity-aware AI assistants for medical data analysis. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. 2020 Apr Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25 - 30, 2020; Honolulu HI USA p. 1-14. [doi: 10.1145/3313831.3376506]
16. Holzinger A, Biemann C, Pattichis C, Kell D. What do we need to build explainable AI systems for the medical domain? arXiv. 2017. URL: https://arxiv.org/pdf/1712.09923.pdf [accessed 2021-12-29]
17. Yang Q, Zimmerman J, Steinfeld A, Carey L, Antaki JF. Investigating the heart pump implant decision process: opportunities for decision support tools to help. ACM Trans Comput Hum Interact 2016 May;2016:4477-4488 [FREE Full text] [doi: 10.1145/2858036.2858373] [Medline: 27833397]
18. Cai CJ, Winter S, Steiner D, Wilcox L, Terry M. "Hello AI": uncovering the onboarding needs of medical practitioners for human-AI collaborative decision-making. Proc ACM Hum-Comput Interact 2019 Nov 07;3(CSCW):1-24. [doi: 10.1145/3359206]
19. Younes M, Hanly PJ. Minimizing interrater variability in staging sleep by use of computer-derived features. J Clin Sleep Med 2016 Oct 15;12(10):1347-1356 [FREE Full text] [doi: 10.5664/jcsm.6186] [Medline: 27448418]
20. Ribera M, Lapedriza A. Can we do better explanations? A proposal of user-centered explainable AI. IUI Workshops. 2019. URL: https://explainablesystems.comp.nus.edu.sg/2019/wp-content/uploads/2019/02/IUI19WS-ExSS2019-12.pdf [accessed 2021-12-29]
21. Barda AJ, Horvat CM, Hochheiser H. A qualitative research framework for the design of user-centered displays of explanations for machine learning model predictions in healthcare. BMC Med Inform Decis Mak 2020 Oct 08;20(1):257 [FREE Full text] [doi: 10.1186/s12911-020-01276-x] [Medline: 33032582]
22. Lim BY, Yang Q, Abdul A, Wang D. Why these explanations? Selecting intelligibility types for explanation goals. IUI Workshops. 2019. URL: https://explainablesystems.comp.nus.edu.sg/2019/wp-content/uploads/2019/02/IUI19WS-ExSS2019-20.pdf [accessed 2021-12-29]
23. Yang Q, Steinfeld A, Zimmerman J. Unremarkable AI: fitting intelligent decision support into critical, clinical decision-making processes. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK p. 1-11. [doi: 10.1145/3290605.3300468]
24. Rasmussen J. Skills, rules, and knowledge; signals, signs, and symbols, and other distinctions in human performance models. IEEE Trans Syst Man Cybern 1983 May;SMC-13(3):257-266. [doi: 10.1109/TSMC.1983.6313160]
25. Ravanelli M, Bengio Y. Speaker recognition from raw waveform with SincNet. In: Proceedings of the IEEE Spoken Language Technology Workshop (SLT). 2018 Presented at: IEEE Spoken Language Technology Workshop (SLT); Dec. 18-21, 2018; Athens, Greece. [doi: 10.1109/slt.2018.8639585]
26. Wang J, Wang Z, Li J, Wu J. Multilevel wavelet decomposition network for interpretable time series analysis. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018 Presented at: KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 19 - 23, 2018; London United Kingdom p. 2437-2446. [doi: 10.1145/3219819.3220060]
27. Lee J, Park J, Kim K, Nam J. Sample-level deep convolutional neural networks for music auto-tagging using raw waveforms. arXiv. 2017. URL: https://arxiv.org/abs/1703.01789 [accessed 2021-12-29]
28. Lee T, Hwang J, Lee H. TRIER: template-guided neural networks for robust and interpretable sleep stage identification from EEG recordings. arXiv. 2020. URL: https://arxiv.org/pdf/2009.05407.pdf [accessed 2021-12-29]

29. Simonyan K, Vedaldi A, Zisserman A. Deep inside convolutional networks: visualising image classification models and saliency maps. arXiv. 2013. URL: https://arxiv.org/pdf/1312.6034.pdf [accessed 2021-12-29]
30. Pons J, Serra X. Randomly weighted CNNs for (music) audio classification. In: Proceedings of the ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2019 Presented at: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); May 12-17, 2019; Brighton, UK. [doi: 10.1109/icassp.2019.8682912]
31. Yosinski J, Clune J, Nguyen A, Fuchs T, Lipson H. Understanding neural networks through deep visualization. arXiv. 2015. URL: https://arxiv.org/abs/1506.06579 [accessed 2021-12-29]
32. Albawi S, Mohammed TA, Al-Zawi S. Understanding of a convolutional neural network. In: Proceedings of the International Conference on Engineering and Technology (ICET). 2017 Presented at: International Conference on Engineering and Technology (ICET); Aug. 21-23, 2017; Antalya, Turkey. [doi: 10.1109/icengtechnol.2017.8308186]
33. Khalighi S, Sousa T, Santos JM, Nunes U. ISRUC-Sleep: a comprehensive public dataset for sleep researchers. Comput Methods Programs Biomed 2016 Feb;124:180-192. [doi: 10.1016/j.cmpb.2015.10.013] [Medline: 26589468]
34. Zhu X, Cimino JJ. Clinicians' evaluation of computer-assisted medication summarization of electronic medical records. Comput Biol Med 2015 Apr;59:221-231 [FREE Full text] [doi: 10.1016/j.compbiomed.2013.12.006] [Medline: 24393492]
35. Cha KH, Hadjiiski LM, Cohan RH, Chan H, Caoili EM, Davenport MS, et al. Diagnostic accuracy of CT for prediction of bladder cancer treatment response with and without computerized decision support. Acad Radiol 2019 Sep;26(9):1137-1145 [FREE Full text] [doi: 10.1016/j.acra.2018.10.010] [Medline: 30424999]
36. Kostopoulou O, Rosen A, Round T, Wright E, Douiri A, Delaney B. Early diagnostic suggestions improve accuracy of GPs: a randomised controlled trial using computer-simulated patients. Br J Gen Pract 2015 Jan;65(630):49-54 [FREE Full text] [doi: 10.3399/bjgp15X683161] [Medline: 25548316]
37. Verdoorn S, Kwint HF, Hoogland P, Gussekloo J, Bouvy ML. Drug-related problems identified during medication review before and after the introduction of a clinical decision support system. J Clin Pharm Ther 2018 Apr;43(2):224-231. [doi: 10.1111/jcpt.12637] [Medline: 28971492]
38. Whitney CW, Gottlieb DJ, Redline S, Norman RG, Dodge RR, Shahar E, et al. Reliability of scoring respiratory disturbance indices and sleep staging. Sleep 1998 Nov 01;21(7):749-757. [doi: 10.1093/sleep/21.7.749] [Medline: 11286351]
39. Cohen J. A coefficient of agreement for nominal scales. Educ Psychol Meas 2016 Jul 02;20(1):37-46. [doi: 10.1177/001316446002000104]
40. Hoff KA, Bashir M. Trust in automation: integrating empirical evidence on factors that influence trust. Hum Factors 2015 May;57(3):407-434. [doi: 10.1177/0018720814547570] [Medline: 25875432]
41. Wilcoxon F. Individual comparisons by ranking methods. In: Breakthroughs in Statistics. New York: Springer; 1992:196-202.
42. Jiang Y, Lee MT, He X, Rosner B, Yan J. Wilcoxon rank-based tests for clustered data with R package clusrank. J Stat Soft 2020;96(6):1-26. [doi: 10.18637/jss.v096.i06]
43. Rosner B, Glynn RJ, Lee MT. The Wilcoxon signed rank test for paired comparisons of clustered data. Biometrics 2006 Mar;62(1):185-192. [doi: 10.1111/j.1541-0420.2005.00389.x] [Medline: 16542245]
44. Datta S, Satten GA. A signed-rank test for clustered data. Biometrics 2008 Jun;64(2):501-507. [doi: 10.1111/j.1541-0420.2007.00923.x] [Medline: 17970820]
45. Kerby DS. The simple difference formula: an approach to teaching nonparametric correlation. Compreh Psychol 2014 Feb 14;3. [doi: 10.2466/11.it.3.1]
46. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
47. Cai C, Reif E, Hegde N, Hipp J, Kim B, Smilkov D, et al. Human-centered tools for coping with imperfect algorithms during medical decision-making. In: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. 2019 Presented at: CHI '19: CHI Conference on Human Factors in Computing Systems; May 4 - 9, 2019; Glasgow Scotland UK p. 1-14. [doi: 10.1145/3290605.3300234]
48. Heer J. Agency plus automation: designing artificial intelligence into interactive systems. Proc Natl Acad Sci U S A 2019 Feb 05;116(6):1844-1850 [FREE Full text] [doi: 10.1073/pnas.1807184115] [Medline: 30718389]
49. Combrisson E, Vallat R, Eichenlaub J, O'Reilly C, Lajnef T, Guillot A, et al. Sleep: an open-source python software for visualization, analysis, and staging of sleep data. Front Neuroinform 2017;11:60 [FREE Full text] [doi: 10.3389/fninf.2017.00060] [Medline: 28983246]
50. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]
51. Treisman AM, Gelade G. A feature-integration theory of attention. Cogn Psychol 1980 Jan;12(1):97-136. [doi: 10.1016/0010-0285(80)90005-5] [Medline: 7351125]
52. Ashcraft M, Radvansky G. Cognition. 6th Ed. Upper Saddle River, NJ: Pearson Education; 2014.
53. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. J Exp Psychol Gen 2015 Feb;144(1):114-126. [doi: 10.1037/xge0000033] [Medline: 25401381]
54. Lyell D, Magrabi F, Coiera E. Reduced verification of medication alerts increases prescribing errors. Appl Clin Inform 2019 Jan;10(1):66-76 [FREE Full text] [doi: 10.1055/s-0038-1677009] [Medline: 30699458]

55. Guo C, Pleiss G, Sun Y, Weinberger K. On calibration of modern neural networks. arXiv. 2017. URL: https://arxiv.org/pdf/1706.04599.pdf [accessed 2021-12-29]
56. Parekh A, Selesnick IW, Rapoport DM, Ayappa I. Detection of K-complexes and sleep spindles (DETOKS) using sparse optimization. J Neurosci Methods 2015 Aug 15;251:37-46. [doi: 10.1016/j.jneumeth.2015.04.006] [Medline: 25956566]
57. Bremer G, Smith JR, Karacan I. Automatic detection of the K-complex in sleep electroencephalograms. IEEE Trans Biomed Eng 1970 Oct;17(4):314-323. [doi: 10.1109/tbme.1970.4502759] [Medline: 5518827]
58. Warby SC, Wendt SL, Welinder P, Munk EG, Carrillo O, Sorensen HB, et al. Sleep-spindle detection: crowdsourcing and evaluating performance of experts, non-experts and automated methods. Nat Methods 2014 Apr;11(4):385-392 [FREE Full text] [doi: 10.1038/nmeth.2855] [Medline: 24562424]
59. Xie Y, Chen M, Kao D, Gao G, Chen X. CheXplain: enabling physicians to explore and understand data-driven, AI-enabled medical imaging analysis. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: CHI '20: CHI Conference on Human Factors in Computing Systems; April 25 - 30, 2020; Honolulu HI USA p. 1-13. [doi: 10.1145/3313831.3376807]
60. Hart SG. Nasa-Task Load Index (NASA-TLX); 20 years later. Proc Hum Factors Ergon Soc Annu Meet 2016 Nov 05;50(9):904-908 [FREE Full text] [doi: 10.1177/154193120605000909]

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
EEG: electroencephalogram
EMG: electromyogram
EOG: electrooculogram
REM: rapid eye movement
TRIER: The Template-Guided Neural Networks for Robust and Interpretable Sleep Stage Identification from EEG Recordings

Edited by A Mavragani; submitted 09.03.21; peer-reviewed by D Lyell, L Grepo; comments to author 05.05.21; revised version received 30.06.21; accepted 01.12.21; published 19.01.22.

Please cite as:
Hwang J, Lee T, Lee H, Byun S
A Clinical Decision Support System for Sleep Staging Tasks With Explanations From Artificial Intelligence: User-Centered Design and Evaluation Study
J Med Internet Res 2022;24(1):e28659
URL: https://www.jmir.org/2022/1/e28659
doi:10.2196/28659
PMID:35044311

Original Paper

The Development and Validation of Simplified Machine Learning Algorithms to Predict Prognosis of Hospitalized Patients With COVID-19: Multicenter, Retrospective Study

Fang He[1,2], BChE, PhD; John H Page[3], MD, SCD; Kerry R Weinberg[2], BSc, MBA, MSci; Anirban Mishra[2], BSc, MSc

[1]Amgen Inc, Center for Observational Research, South San Francisco, CA, United States

[2]Amgen Inc, Digital Health & Innovation, Thousand Oaks, CA, United States

[3]Amgen Inc, Center for Observational Research, Thousand Oaks, CA, United States

Corresponding Author:
Fang He, BChE, PhD
Amgen Inc
Center for Observational Research
1120 Veterans Boulevard
South San Francisco, CA, 94080
United States
Phone: 1 925 216 1658
Email: fhe01@amgen.com

Abstract

Background: The current COVID-19 pandemic is unprecedented; under resource-constrained settings, predictive algorithms can help to stratify disease severity, alerting physicians of high-risk patients; however, there are only few risk scores derived from a substantially large electronic health record (EHR) data set, using simplified predictors as input.

Objective: The objectives of this study were to develop and validate simplified machine learning algorithms that predict COVID-19 adverse outcomes; to evaluate the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and calibration of the algorithms; and to derive clinically meaningful thresholds.

Methods: We performed machine learning model development and validation via a cohort study using multicenter, patient-level, longitudinal EHRs from the Optum COVID-19 database that provides anonymized, longitudinal EHR from across the United States. The models were developed based on clinical characteristics to predict 28-day in-hospital mortality, intensive care unit (ICU) admission, respiratory failure, and mechanical ventilator usages at inpatient setting. Data from patients who were admitted from February 1, 2020, to September 7, 2020, were randomly sampled into development, validation, and test data sets; data collected from September 7, 2020, to November 15, 2020, were reserved as the postdevelopment prospective test data set.

Results: Of the 3.7 million patients in the analysis, 585,867 patients were diagnosed or tested positive for SARS-CoV-2, and 50,703 adult patients were hospitalized with COVID-19 between February 1 and November 15, 2020. Among the study cohort (n=50,703), there were 6204 deaths, 9564 ICU admissions, 6478 mechanically ventilated or EMCO patients, and 25,169 patients developed acute respiratory distress syndrome or respiratory failure within 28 days since hospital admission. The algorithms demonstrated high accuracy (AUC 0.89, 95% CI 0.89-0.89 on the test data set [n=10,752]), consistent prediction through the second wave of the pandemic from September to November (AUC 0.85, 95% CI 0.85-0.86) on the postdevelopment prospective test data set [n=14,863], great clinical relevance, and utility. Besides, a comprehensive set of 386 input covariates from baseline or at admission were included in the analysis; the end-to-end pipeline automates feature selection and model development. The parsimonious model with only 10 input predictors produced comparably accurate predictions; these 10 predictors (age, blood urea nitrogen, SpO_2, systolic and diastolic blood pressures, respiration rate, pulse, temperature, albumin, and major cognitive disorder excluding stroke) are commonly measured and concordant with recognized risk factors for COVID-19.

Conclusions: The systematic approach and rigorous validation demonstrate consistent model performance to predict even beyond the period of data collection, with satisfactory discriminatory power and great clinical utility. Overall, the study offers an accurate, validated, and reliable prediction model based on only 10 clinical features as a prognostic tool to stratifying patients with COVID-19 into intermediate-, high-, and very high-risk groups. This simple predictive tool is shared with a wider health

care community, to enable service as an early warning system to alert physicians of possible high-risk patients, or as a resource triaging tool to optimize health care resources.

(J Med Internet Res 2022;24(1):e31549) doi:10.2196/31549

KEYWORDS

COVID-19; predictive algorithm; prognostic model; machine learning

Introduction

The COVID-19 pandemic has impacted more than 200 countries, claimed more than 3 million lives, presenting an urgent threat to global health. Under resource-constrained settings, a validated model using large-scale real-world data to predict COVID-19 prognosis can rapidly identify the individuals who are at risk of COVID-19 adverse outcomes and mortality, so they could benefit from early interventions.

Several studies have derived prognostic predictors for COVID-19; however, currently there are only few COVID-19 risk calculation tools with simplified predictors for stratification that leverage on a substantially large US electronic health record (EHR) data set of statistically meaningful size [1,2]. The Acute Physiology and Chronic Health Evaluation (APACHE) II score [3] has been widely used to predict in-hospital mortality, and has been found to predict mortality in patients with COVID-19, outperforming Sequential Organ Failure Assessment (SOFA) [4] and CURB-65 [5] scores in a retrospective study of 154 patients in China [6]. COVID-GRAM [2] is a web-based calculator to estimate the occurrence of ICU admission, mechanical ventilation, or death in hospitalized patients with COVID-19; it has been validated in a study of nearly 1600 patients in China. The Coronavirus Clinical Characterization Consortium (4C) Mortality Score [1] developed by the International Severe Acute Respiratory and Emerging Infections Consortium (ISARIC) World Health Organization (WHO) Clinical Characterisation Protocol UK (CCP-UK) study is a risk stratification tool to predict in-hospital mortality by categorizing patients at low, intermediate, high, or very high risk of death. Separately, an accurate, machine learning–based COVID-19 mortality prediction model has been developed based on data from the Mount Sinai Health System; however, its validation data set is limited in size [7].

The objective of this paper is to develop and validate simplified and parsimonious predictive algorithms, leveraging large size, near real-time real-world data as a risk stratification methodology to identify patients who are at heightened risk of (1) mortality; (2) ICU admission; (3) composite of invasive mechanical ventilation/extracorporeal membrane oxygenation (ECMO); (4) composite of acute respiratory distress syndrome (ARDS)/respiratory failure, which can be easily integrated into the hospital electronic medical record system as a risk stratification and triaging tool.

Methods

Data Source

This is a retrospective observational cohort analysis of multicenter, longitudinal, anonymized patient-level data from the Optum EHR COVID-19 database. It includes demographics, insurance status, medication prescription, vital signs, coded diagnoses, procedures, laboratory results, visits, encounters, and providers. Currently, there are 3,702,050 patients in the data release dated January 27, 2021. As deidentified data are used for the study, it was exempt from Institutional Review Board approval.

Study Period

The study period was from February 1, 2020, to January 27, 2021. A baseline of up to 1 year prior to and including index date was used for assessment of demographics, lifestyle factors, and comorbidity at baseline. Patients were followed up to 28 days from admission, unless they were censored by in-hospital mortality or discharged.

Participants

Study cohort consists of patients hospitalized with COVID-19 aged 18 and older, with a confirmed diagnosis or positive test of COVID-19 infection. A COVID-19 diagnosis was defined as the first occurrence on or after February 1, 2020, of any of the following: (1) positive result from SARS-CoV-2 viral RNA or antigen tests; (2) International Classification of Diseases, Tenth Revision, Clinical Modification (ICD-10-CM) diagnosis codes U07.1 (COVID-19, virus identified), J12.81 (pneumonia due to SARS-associated coronavirus), J12.89 (other viral pneumonia), or J80 (ARDS); and (3) ICD-10-CM code B97.29 (other coronavirus as the cause of disease) or B34.20 (coronavirus infection, unspecified) occurring on or before April 30, 2020. The expanded diagnosis code list, beyond COVID-19–specific diagnosis code (U07.1), was used because U07.1 was either unavailable (pre-April 2020) or was being implemented (April 2020), resulting in the use of alternative codes for COVID-19 in early pandemic. Other codes such as J20.3 (acute bronchitis due to coxsackievirus) were excluded due to very few uses (<10 patients) in the study period.

Patients were excluded for any of the following: (1) missing age or sex; (2) under the age of 18; (3) diagnosis or procedure codes for labor and delivery during hospitalization; (4) diagnosis codes for trauma, injury, fracture, or poisoning during the first 2 days of hospitalization; (5) admitted to hospital more than 10 days prior to COVID-19 diagnosis or 28 days after COVID-19 diagnosis; (6) diagnosed or admitted to hospital after November 16, 2020, therefore with less than 10 weeks between their first COVID-19 diagnosis date or hospital admission date, and the last database refresh date (January 27, 2021; Figure 1). Additional sensitivity analysis was conducted between the final study cohort (n=50,703) and patients who tested positive for SARS-CoV-2 (n=38,277), a subset of the former.

Sampling

In the final cohort that satisfied the study criteria (n=50,703), data from patients with an index date prior to September 7, 2020, were referred to as model development data set (n=35,840), which was randomly sampled without replacement using 28-day in-hospital mortality as stratification factor into 40% training data set (n=14,336), 30% validation data set (n=10,752) for hyperparameter tuning and threshold calculation, and 30% test data set (n=10,752). The sampling ratio is determined such that the validation or test data set alone can satisfy the sample size requirement; the minimum sample size is estimated to be 8605, assuming a predetermined sensitivity of 0.7 and the prevalence of all-cause mortality of 15% with 95% CI and maximum marginal error of estimate of 2.5% [8]. Furthermore, an independent validation consisting of patients with index date from September 7 to November 15, 2020, was referred to as postdevelopment prospective test data set (n=14,863).

Index Date

The index date was defined as hospital admission date.

Sample Size

The initial anonymized data for 3,702,050 patients from 885,677 providers and 2465 delivery networks for the study period February 1, 2020, to January 27, 2021, were transferred from Optum, among which 585,867 patients were diagnosed or tested positive for SARS-CoV-2 infection.

Outcome

The outcomes were 28-day in-hospital (1) all-cause mortality; (2) ICU admission; (3) composite of invasive mechanical ventilation or ECMO; (4) composite of ARDS and respiratory failure. These were assessed as dichotomous outcomes and individually modeled. Outcome-specific exclusions were applied as appropriate to include only incident outcomes.

Covariates

A total of 386 study covariates (with a minimum 70% [35,493/50,703 patients] coverage among study cohort) consisting of patients' baseline demographics (age, sex, census division, insurance status, race, ethnicity), lifestyle factors (smoking status, BMI), comorbidities (including atrial fibrillation cancer history, cerebrovascular disease, chronic kidney disease stage I-V, chronic obstructive pulmonary disease, coronary artery disease, Type I/II diabetes mellitus, HIV, stroke, etc.), baseline medication (including antidiabetics, anticoagulants, antihypertensives, antiplatelets, steroids, etc.) within 12 months prior to index date, vital signs (blood pressures, heart rate, pulse, respiration rate, temperature), laboratory values (including albumin, alanine transaminase, aspartate aminotransferase, total bilirubin, B-type natriuretic peptide, blood urea nitrogen (BUN), chloride, creatinine, C-reactive protein, D-dimer, fibrinogen, hemoglobin, lymphocyte, monocyte, neutrophil, oxygen saturation platelet count, arterial blood pH, etc.), and treatment (including diuretics, disease-modifying antirheumatic drugs, steroids, etc.) administered on the day of hospital admission were included in the analysis. Concretely, baseline medication, comorbidity, and postadmission treatment were expressed as dichotomous variables; categorical variables were converted to dummy variables; numerical variables were used without standardization, unless when fitting to penalized (Lasso or Ridge) logistic regression models, while numerical covariates were normalized using a min–max standardization to speed up convergence.

Missing Data

One of the challenges of working with real-world data is the missing covariates. Assuming covariates are missing at random, multiple imputation by chained equations via random forest [9] was used to impute covariates with missing values. Ten complete data sets each with 10 iterations were imputed with predictive mean matching using available covariates while excluding the outcome variables. The prediction performances of sparsity-aware models (XGB [10]) between imputed and nonimputed data set were compared in the sensitivity analysis.

Given the intention to develop an algorithm of great relevance to as many patients as possible, we have restricted the model input to covariates with a minimum of 70% coverage in the study cohort. Overall, the proportion of missing values among the vital and laboratory variables ranges from 10.44% (5295/50,703) to 99.36% (50,381/50,703) out of 50,703 patients; 45 of 431 variables were not included as input to the model due to more than 30% (15,211/50,703) of missingness (Multimedia Appendices 1 and 2). Sensitivity analysis was conducted to evaluate whether inclusion of additional covariates with higher degrees of missingness (ie, varying the cutoffs from 10% to 90%) aids in improving model performance, though it may increase the sensitivity of the models to biases due to nonignorable missingness in the data.

Model Development

We have applied a systematic approach to model development and validation. A framework of 6 machine learning algorithms (XGB [10,11], penalized logistic regression [12,13] with Lasso [14] or Ridge loss [11], random forest [11,15,16], decision tree [17], and LightGBM [18]) has been adopted to develop interpretable models to predict the prognosis of COVID-19.

In the preliminary analysis, the most performant algorithm was selected from the candidate algorithms; prior to model training, hyperparameter optimization via grid search, ranging from 96 to 243 folds, was performed on 6 candidate algorithms individually for full and simplified models. The full model uses all the available 386 input features after extraction and transformation in preliminary analysis, while the simplified model recursively eliminates the aforementioned input to yield a maximum of 20 variables [19]. The algorithm with best performance (area under the receiver operating characteristic curve [AUC], Brier score [20], and calibration [21]) on the test data set was selected for the final analysis.

In the final analysis, model input is further iteratively reduced to a maximum of 5 variables with a step size of 1; 100 individual runs were performed at each step, with retuned model parameters every 5 steps. The selected features were pooled and plotted in frequency heatmap with the corresponding AUC.

The model performance is evaluated against outcome variables in the test and postdevelopment prospective test data sets via AUC, Brier score, and calibration curve. The 95% CIs for AUC and Brier score were calculated based on percentiles from bootstrapped resampling with replacement (bootstrap sample size = 2000) without bias correction or acceleration [22]. The calibration curves (number of discretized bins = 10) were plotted for all the runs.

Model Validation

Rigorous validation analysis was performed to ensure robustness and reliability of the predictions. Both full and simplified models of 6 candidate algorithms were trained and validated during the model development phase with data from February 1 to September 6, where the test data set was held out from model training and used solely for reporting the performance. Furthermore, the model has been additionally validated externally, using the postdevelopment prospective test data set collected from September 7 to November 15, 2020, to demonstrate consistent model performance through the subsequent wave of the pandemic. Model discrimination was performed on the imputed test data set by assessing AUC on the stratified analysis by sex, age, and racial groups.

Model Benchmark

The performance of the risk prediction models has been benchmarked to (1) the baseline model and (2) published COVID-19 prognostic scores. The baseline model was developed using XGB with optimized hyperparameters on age and sex only. Evaluation metrics including AUC, sensitivity, specificity, and decision curve analysis were assessed to compare the performance and utility of prognostic scores (APACHE II [3,23], CURB-65 [5,24,25], E-CURB [26], The National Early Warning [NEWS] 2 score [25,27-29], Respiratory Rate-Oxygenation [ROX] index [29,30], ISARIC 4C mortality score [1,25]). AUC is reported based on complete case data from test and postdevelopment prospective test data sets, and no imputation was performed.

Predictors

Feature importance is ranked by Shapley values [31] from test and postdevelopment prospective test data sets in the SHapley Additive exPlanations (SHAP) summary plot. Shapley value calculates fair contribution and the extent of predictors toward the model output [32]. It measures feature importance by the magnitude and the direction of contributions. The dependence between model prediction and age is plotted with age on the x-axis and its impact on prediction represented by Shapley value on the y-axis for every patient, colored by the magnitude of a second feature (BUN, respiration rate, pulse, lymphocyte count) individually.

Receiver Operating Characteristic Curve Analysis

We adopted two approaches in determining the optimal threshold on the receiver operating characteristic curve. Assuming the sensitivity and specificity were weighted equally without ethical, cost, and prevalence constraints, the optimal cutoff is at the location where the Youden index (sum of specificity and sensitivity – 1) is maximized at the test data set [33-36]. This approach relies solely on the predictive accuracy of a model, and consequences of the predictions (ie, cost of false positives and false negatives) are not considered. In the second approach, clinical utility–based decision theory was used in developing a cost-sensitive threshold, where it builds in disease prevalence and costs of false positive and false negatives of specific diagnostic scenario [33,37].

Decision curve analysis assists in clinical judgment and comparison about the relative value of benefits associated with the use of a clinical prediction tool [38,39]. The standardized net benefit of full model, simplified model (with 10 input variables), and selected benchmark prognostic scores was calculated and plotted across probabilities. The benchmark models that use point scores were calibrated to test data prior to decision curve analysis.

Results

Patient Characteristics

Figure 1 shows patient attrition flowchart, and the workflow of model development and validation is in Figure 2. Patients' baseline and clinical characteristics at admission are summarized in Table 1. Validation and test data sets are largely homogeneous to the training data set; however, the postdevelopment prospective test data set that was collected later in the pandemic from September to November presents more differences in geographic locations (a decline in the proportion of patients in Middle Atlantic from 22.69% [8132/35,840] to 8.68% [1290/14,863] after September 7, and an increase in West North Central from 9.83% [3523/35,840] to 24.18% [3594/14,863]) and racial distribution (the proportion of White increased from 53.87% [19,308/35,840] to 72.88% [10,832/14,863]). However, the overall mortality remains consistent. Hypertension (57.55% [29,179/50,703]), obesity (47.51% [24,089/50,703]), diabetes mellitus (34.44% [17,461/50,703]), chronic kidney disease (19.79% [10,033/50,703]), and coronary artery disease (17.74% [8996/50,703]) were the common comorbidities among the cohort.

The study cohort is defined as hospitalized adult patients with COVID-19 who were either diagnosed with relevant diagnosis codes or tested positive for SARS-CoV-2 viral RNA or antigen tests. In the subgroup analysis of patients who tested positive for SARS-CoV-2, model performances of these 2 groups (ie, overall cohort and tested positive subgroup) were largely similar with less than 1% difference in AUC across all outcomes for full and simplified models (Multimedia Appendix 3).

Figure 1. Patient attrition diagram. ^With relevant COVID-19 diagnosis codes or tested positive for SARS-CoV-2. *Non-exclusive critera: overlapping was allowed.

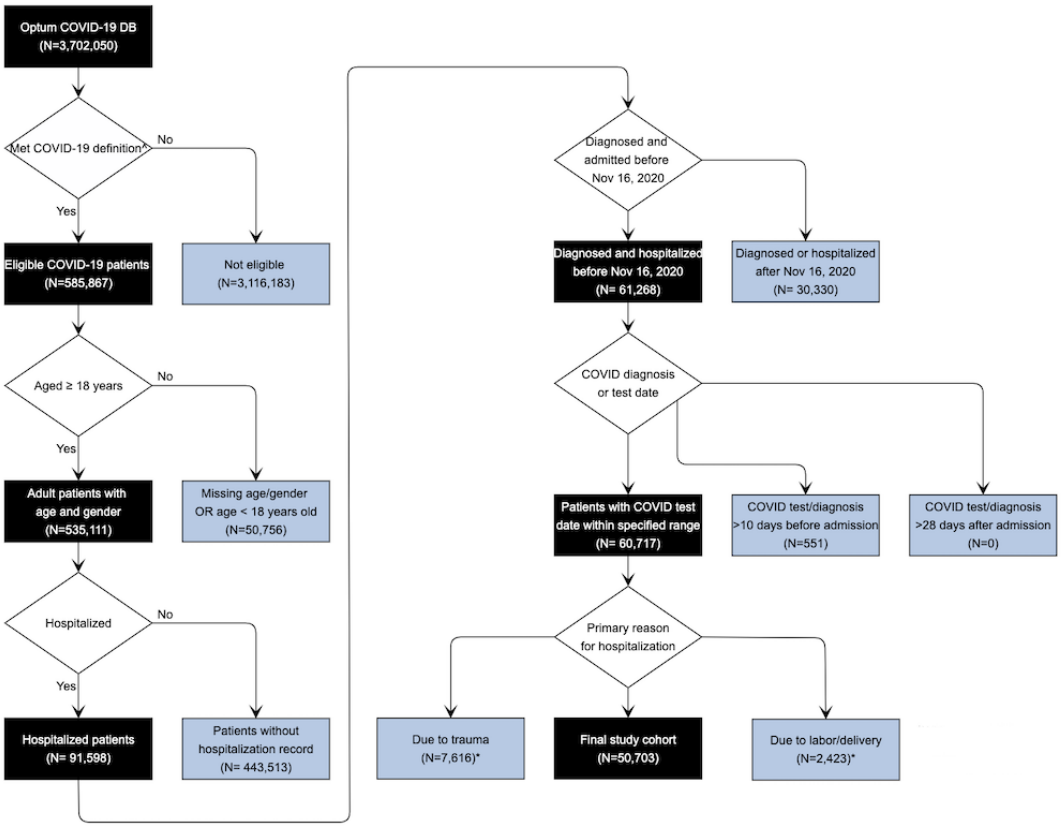


Figure 2. Model development and validation framework including data sampling and corresponding sensitivity analyses.

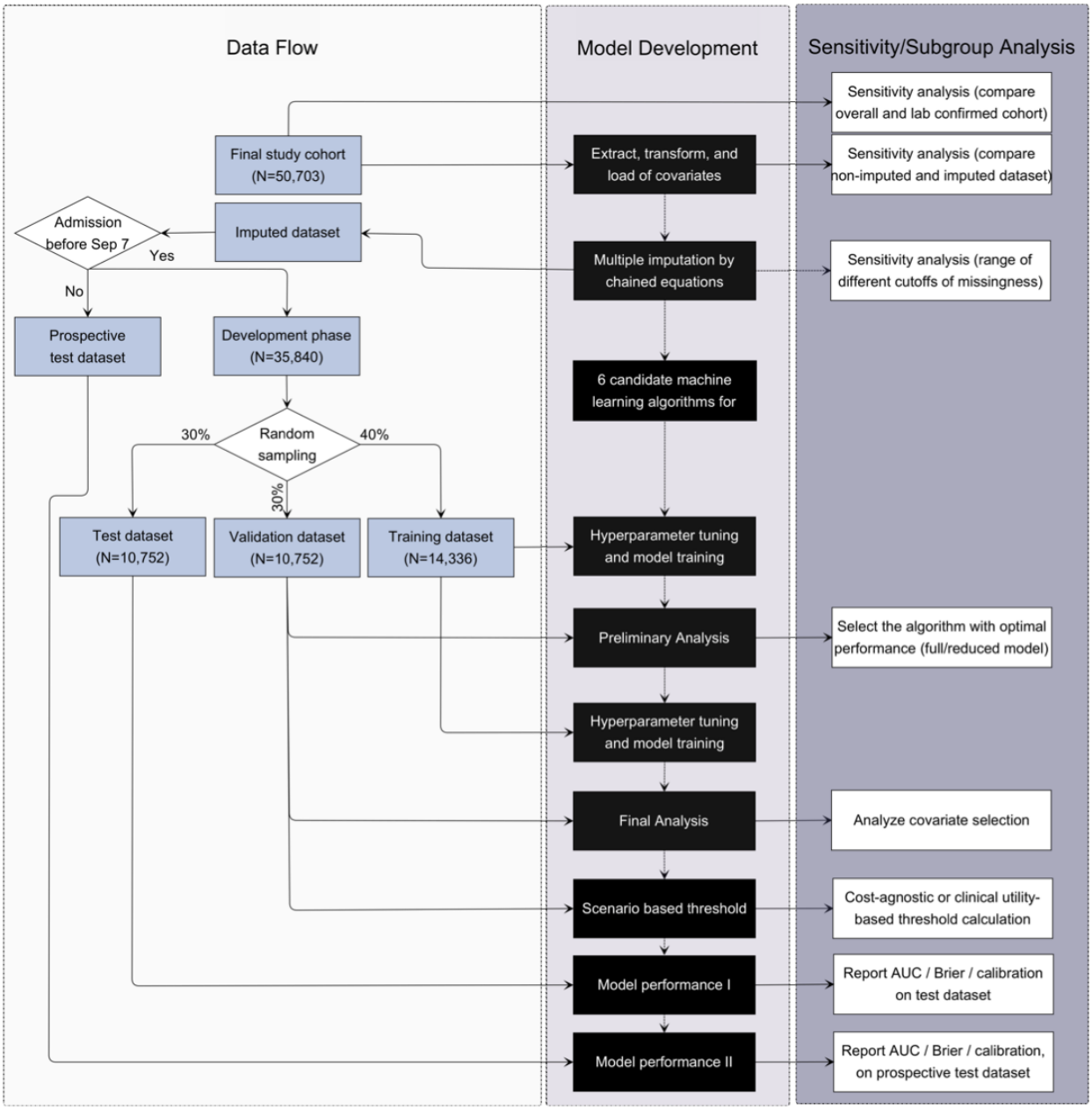


Table 1. Demographic and clinical characteristics of hospitalized patients with COVID-19 at baseline and admission.

Characteristic	Training data set (n=14,336)	Validation data set (n=10,752)	Test data set (n=10,752)	Prospective test data set (n=14,863)
Mean (SD) age at baseline, years	60.9 (17.2)	60.9 (17.2)	60.8 (17.1)	63.8 (16.8)
Distribution, n (%)				
18-34	1231 (8.59)	920 (8.56)	911 (8.47)	1015 (6.83)
35-49	2383 (16.62)	1780 (16.56)	1840 (17.11)	1893 (12.74)
50-64	4337 (30.25)	3193 (29.70)	3293 (30.63)	4110 (27.65)
65-74	2922 (20.38)	2296 (21.35)	2141 (19.91)	3325 (22.37)
75-84	2165 (15.10)	1589 (14.78)	1606 (14.94)	2943 (19.80)
85+	1298 (9.05)	974 (9.06)	961 (8.94)	1577 (10.61)
Sex at baseline, n (%)				
Male	7473 (52.13)	5619 (52.26)	5629 (52.35)	7645 (51.44)
Female	6863 (47.87)	5133 (47.74)	5123 (47.65)	7218 (48.56)
Race at baseline, n (%)				
African American	3466 (24.18)	2669 (24.82)	2668 (24.81)	1867 (12.56)
Asian	368 (2.57)	268 (2.49)	276 (2.57)	216 (1.45)
White	7779 (54.26)	5734 (53.33)	5795 (53.90)	10,832 (72.88)
Other/Unknown	2723 (18.99)	2081 (19.35)	2013 (18.72)	1948 (13.11)
Census division at baseline, n (%)				
East North Central	3778 (26.35)	2942 (27.36)	2908 (27.05)	4174 (28.08)
East South Central	1010 (7.05)	708 (6.58)	754 (7.01)	1205 (8.11)
Middle Atlantic	3221 (22.47)	2488 (23.14)	2423 (22.54)	1290 (8.68)
Mountain	496 (3.46)	355 (3.30)	363 (3.38)	923 (6.21)
New England	1042 (7.27)	705 (6.56)	763 (7.10)	769 (5.17)
Pacific	475 (3.31)	331 (3.08)	317 (2.95)	345 (2.32)
South Atlantic/West South Central	2454 (17.12)	1802 (16.76)	1810 (16.83)	2120 (14.26)
West North Central	1396 (9.74)	1067 (9.92)	1060 (9.86)	3594 (24.18)
Other/Unknown	464 (3.24)	354 (3.29)	354 (3.29)	443 (2.98)
BMI at baseline (kg/m^2), mean (SD)	31.0 (8.5)	30.9 (8.3)	31.2 (8.6)	31.6 (8.7)
Distribution, n (%)				
Underweight	352 (2.46)	235 (2.19)	221 (2.06)	304 (2.05)
Healthy weight	2526 (17.62)	1873 (17.42)	1833 (17.05)	2283 (15.36)
Overweight	3697 (25.79)	2878 (26.77)	2838 (26.40)	3679 (24.75)
Obese	3041 (21.21)	2247 (20.90)	2344 (21.80)	3228 (21.72)
Morbidly obese	3679 (25.66)	2739 (25.47)	2742 (25.50)	4069 (27.38)
Unknown	1041 (7.26)	780 (7.25)	774 (7.20)	1300 (8.75)
Comorbidity at baseline[a], n (%)				
Cerebrovascular disease	676 (4.72)	502 (4.67)	501 (4.66)	894 (6.01)
Chronic kidney disease	2808 (19.59)	2058 (19.14)	2040 (18.97)	3127 (21.04)
Congestive heart failure	2137 (14.91)	1534 (14.27)	1553 (14.44)	2369 (15.94)
Coronary artery disease	2430 (16.95)	1797 (16.71)	1800 (16.74)	2969 (19.98)
Diabetes mellitus	4831 (33.70)	3636 (33.82)	3586 (33.35)	5408 (36.39)
Hypertension	8173 (57.01)	6091 (56.65)	6063 (56.39)	8852 (59.56)

Characteristic	Training data set (n=14,336)	Validation data set (n=10,752)	Test data set (n=10,752)	Prospective test data set (n=14,863)
Solid tumor	830 (5.79)	606 (5.64)	619 (5.76)	1052 (7.08)
Transplant history	28 (0.20)	16 (0.15)	20 (0.19)	12 (0.08)
28-day outcomes, n (%)				
All-cause mortality	1769 (12.34)	1326 (12.33)	1327 (12.34)	1782 (11.99)
Intensive care unit admission	2813 (19.62)	2181 (20.28)	2148 (19.98)	2422 (16.30)
Acute respiratory distress syndrome (respiratory failure)	7276 (50.75)	5500 (51.15)	5384 (50.07)	7009 (47.16)
Extracorporeal membrane oxygenation (mechanical ventilation)	1962 (13.69)	1535 (14.28)	1498 (13.93)	1483 (9.98)
Vitals at admission, median (10th-90th percentile)				
Diastolic blood pressure (mmHg)[b]	73.0 (56.0-90.0)	73.0 (56.0-90.0)	73.0 (56.0-90.0)	73.0 (56.0-90.0)
Systolic blood pressure (mmHg)[b]	125.0 (100.0-154.0)	125.0 (101.0-155.0)	125.0 (101.0-154.0)	128.0 (103.0-159.0)
Pulse (bpm)[b]	85.0 (64.0-110.0)	85.0 (64.0-110.0)	85.0 (64.0-110.0)	81.0 (61.0-107.6)
Respiratory rate (breaths/minute)[b]	19.0 (16.0-28.0)	19.0 (16.0-28.0)	19.0 (16.0-28.0)	18.0 (16.0-25.0)
Temperature (^{o}C)[b]	36.8 (36.3-37.9)	36.8 (36.3-37.9)	36.8 (36.3-37.8)	36.7 (36.2-37.7)
Laboratory values[a] at admission, median (10th percentile-90th percentile)				
Alkaline phosphatase (IU/L)	77.0 (49.0-137.0)	76.0 (49.0-136.0)	76.0 (48.0-135.0)	78.0 (50.0-134.0)
Alanine aminotransferase (IU/L)	28.0 (12.0-79.0)	29.0 (12.0-80.0)	28.0 (12.0-79.0)	27.0 (12.0-68.0)
Aspartate aminotransferase (IU/L)	37.0 (18.0-95.0)	36.0 (18.0-97.0)	36.0 (18.0-95.0)	34.0 (18.0-80.0)
Albumin (g/dL)	3.5 (2.7-4.2)	3.6 (2.7-4.2)	3.6 (2.7-4.2)	3.6 (2.8-4.2)
Anion gap (mEq/L)	12.0 (7.0-17.0)	12.0 (7.0-17.0)	12.0 (7.0-17.0)	12.0 (7.0-16.0)
Blood urea nitrogen (mg/dL)	16.0 (8.0-47.0)	17.0 (8.0-46.0)	16.0 (8.0-47.0)	18.0 (9.0-44.0)
Bicarbonate (mmol/L)	24.0 (19.0-29.0)	24.0 (19.0-29.0)	24.0 (19.0-29.0)	24.0 (19.0-29.0)
Bilirubin total (mg/dL)	0.6 (0.3-1.2)	0.6 (0.3-1.2)	0.6 (0.3-1.2)	0.6 (0.3-1.1)
C-reactive protein (mg/dL)	85.0 (10.3-229.0)	82.2 (11.0-218.0)	82.0 (10.2-220.0)	73.0 (10.0-206.6)
Chloride (mmol/L)	101.0 (94.0-108.0)	101.0 (94.0-108.0)	101.0 (94.0-108.0)	101.0 (94.0-107.0)
Glucose (mg/dL)	120.0 (91.0-242.0)	121.0 (92.0-236.0)	121.0 (92.0-240.6)	122.0 (91.2-244.0)
Hemoglobin (g/dL)	13.2 (10.0-15.5)	13.2 (10.1-15.6)	13.2 (10.2-15.7)	13.2 (10.1-15.6)
Lymphocyte (%)	14.1 (5.4-30.0)	14.8 (5.6-30.7)	14.6 (5.8-30.2)	14.1 (5.3-30.0)
Monocyte (%)	7.1 (3.1-12.9)	7.0 (3.2-12.6)	7.1 (3.2-12.7)	7.8 (3.6-13.1)
Neutrophil (%)	75.8 (57.0-88.0)	75.0 (57.0-88.0)	75.2 (57.0-88.0)	75.0 (57.0-88.0)
Platelet count (x10^9/L)	210.0 (125.0-351.0)	210.0 (127.0-348.0)	211.0 (126.0-351.0)	205.0 (124.0-335.0)
Potassium (mmol/L)	3.9 (3.3-4.8)	3.9 (3.3-4.8)	3.9 (3.3-4.8)	3.9 (3.3-4.7)
Protein total (g/dL)	7.2 (6.2-8.2)	7.2 (6.2-8.2)	7.3 (6.2-8.2)	7.1 (6.2-8.0)
Red cell distribution width coefficient of variation (%)	13.9 (12.4-17.0)	13.8 (12.4-16.9)	13.8 (12.4-17.0)	13.8 (12.4-16.7)
Sodium (mmol/L)	136.0 (130.0-141.0)	136.0 (131.0-142.0)	136.0 (131.0-141.0)	136.0 (131.0-141.0)
Oxygen saturation pulse oximeter (%)	96.0 (91.0-99.0)	96.0 (90.0-99.0)	96.0 (91.0-99.0)	95.0 (90.0-99.0)
Oxygen saturation pulse oximeter[b] (%)	95.0 (87.0-99.0)	95.0 (87.0-99.0)	95.0 (87.0-99.0)	95.0 (87.0-99.0)
Oxygen saturation pulse oximeter[c] (%)	93.0 (84.0-97.0)	93.0 (84.0-97.0)	93.0 (84.0-97.0)	92.0 (83.0-97.0)

Characteristic	Training data set (n=14,336)	Validation data set (n=10,752)	Test data set (n=10,752)	Prospective test data set (n=14,863)
White blood cell count ($x10^9$/L)	7.1 (4.0-14.1)	7.1 (4.0-13.9)	7.0 (4.0-13.8)	6.9 (3.9-13.5)

[a]Non-exhaustive list.

[b]First measurement on the day of hospital admission.

[c]Minimum measurement on the day of hospital admission.

Model Performance

We have adopted a systematic framework of model development, including a variety of tree-, boosting-, and ensemble-based machine learning models, combined with rigorous validation on statistically meaningful sample size. The model performances (AUC and Brier score) on test and prospective test data sets are summarized in Table 2 and Figure 3. AUC is a widely used metric for performance measurement of classification; Brier score is a proper scoring rule, measuring mean squared error between prediction and outcome, impacted by both discrimination and calibration. Calibration of the algorithm is further assessed by plotting the predicted proportion against the observed proportion of outcome in each decile of risk (Figure 4).

Figure 3. Receiver operating characteristics (AUROC) curves on four prediction outcomes in final analysis: (a) all-cause mortality; (b) respiratory failure including ARDS; (c) ICU admission; (d) invasive mechanical ventilation including ECMO. Full model is colored in black, parsimonious model with ten input variables is colored in orange. Solid line represents model performance on test dataset (n=10,752); dashed line represents post-development prospective test dataset (n=14,863). ARDS: acute respiratory distress syndrome. ECMO: extracorporeal membrane oxygenation.

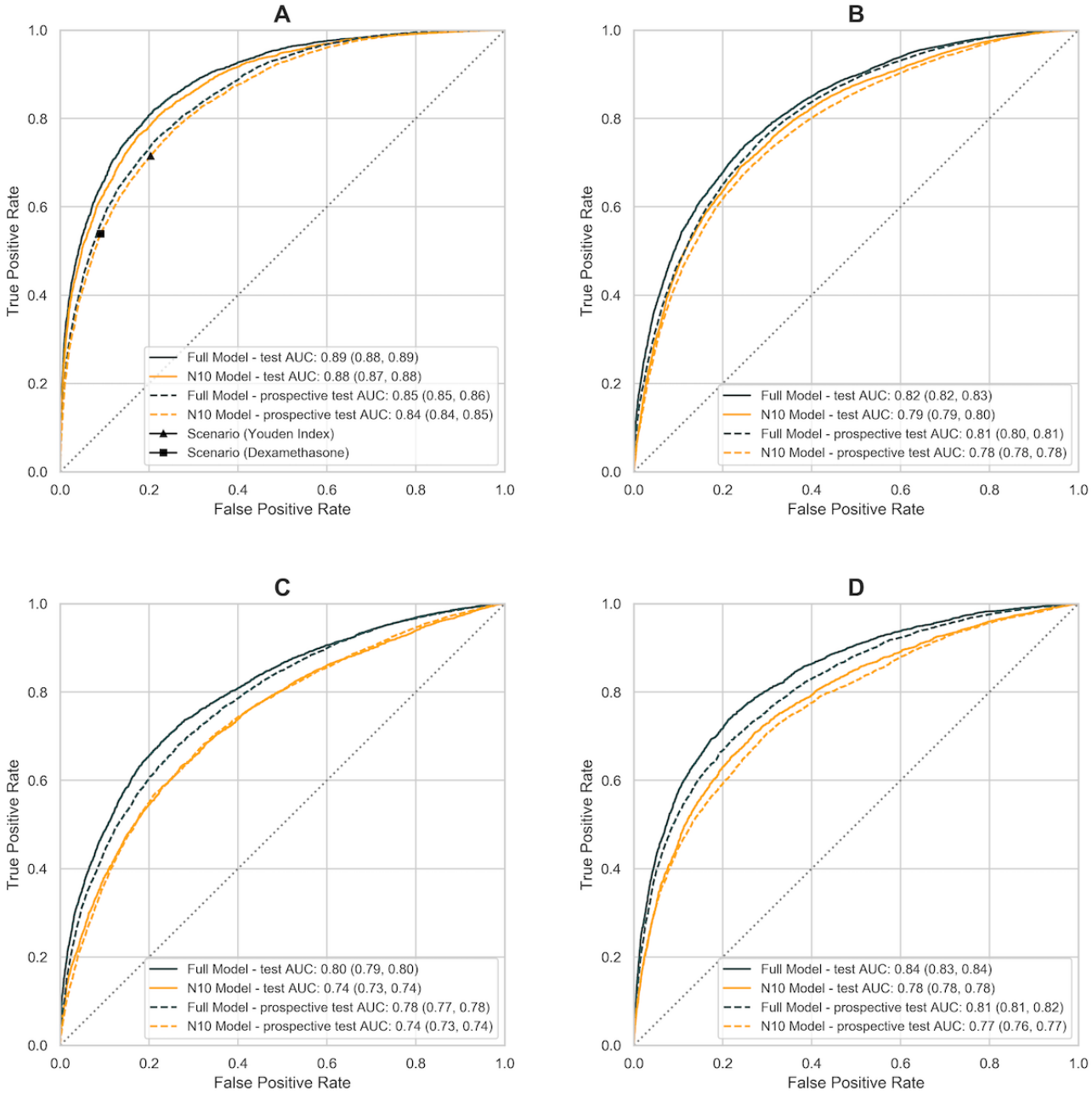


Figure 4. Calibration curve (number of bins = 10) on four prediction outcomes in final analysis: (a) all-cause mortality; (b) respiratory failure including ARDS; (c) ICU admission; (d) invasive mechanical ventilation including ECMO. Full model is colored in black, parsimonious model with ten input variables is colored in orange. Solid line represents calibration on test dataset (n=10,752); dashed line represents calibration on post-development prospective test dataset (n=14,863). ARDS: acute respiratory distress syndrome. ECMO: extracorporeal membrane oxygenation.

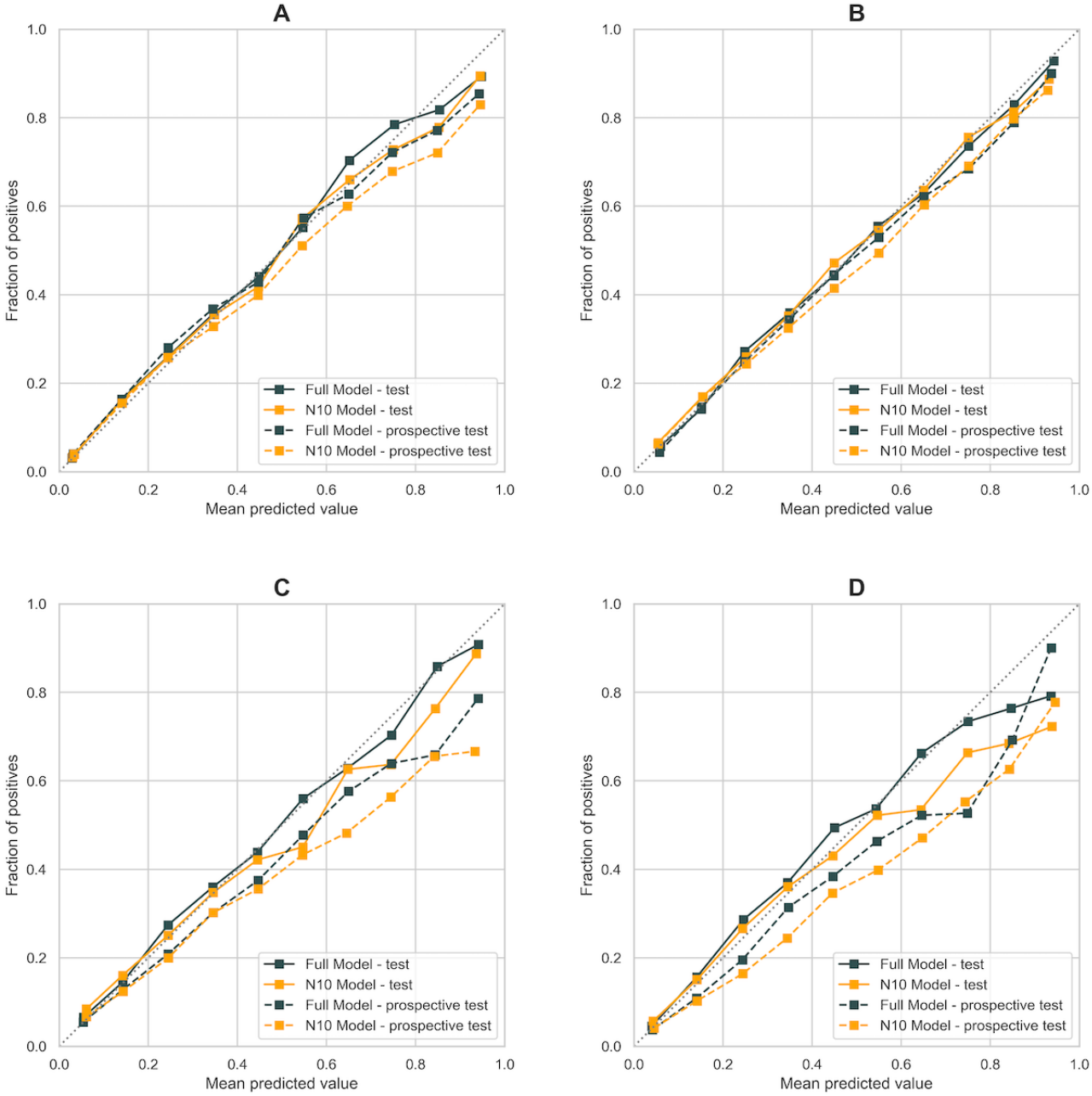


Table 2. Summary of model performances (AUC[a] and Brier Score) on test data set and postdevelopment prospective test data set in the final analysis. The full model uses all the available 210 covariates with less than 30% (15,211/50,703) missingness (excluding postadmission treatment) among the study cohort (n=50,703); the parsimonious N10 model only uses 10 predictors prefiltered from the automatic predictor selection.

Outcome and model	AUC[a] (95% CI)		Brier score (95% CI)	
	Test data set, %	Prospective test data set, %	Test data set	Prospective test data set
All-cause mortality				
Full model	88.7 (88.4-89.0)	85.4 (85.1-85.7)	0.071 (0.070-0.072)	0.079 (0.078-0.080)
N10 model	87.6 (87.2-87.9)	84.3 (84.0-84.6)	0.074 (0.073-0.075)	0.081 (0.080-0.081)
Intensive care unit admission				
Full model	79.7 (79.4-80.1)	77.7 (77.3-78.0)	0.123 (0.122-0.124)	0.115 (0.114-0.115)
N10 model	73.6 (73.2-74.0)	73.5 (73.2-73.9)	0.138 (0.137-0.139)	0.123 (0.122-0.124)
Respiratory failure[b]				
Full model	82.3 (82.0-82.5)	80.7 (80.5-80.9)	0.172 (0.171-0.173)	0.180 (0.179-0.181)
N10 model	79.5 (79.2-79.7)	78.1 (77.9-78.3)	0.185 (0.184-0.186)	0.192 (0.191-0.193)
Mechanical ventilation[c]				
Full model	83.6 (83.3-84.0)	81.1 (80.8-81.5)	0.090 (0.089-0.091)	0.074 (0.074-0.075)
N10 model	78.1 (77.7-78.5)	76.6 (76.2-77.1)	0.101 (0.100-0.101)	0.081 (0.081-0.082)

[a]AUC: area under the receiver operating characteristic curve.

[b]Refers to composite of respiratory failure and acute respiratory distress syndrome.

[c]Refers to composite of invasive mechanical ventilation and extracorporeal membrane oxygenation.

The model predicts 28-day in-hospital mortality accurately (AUC 0.88, 95% CI 0.87-0.88 on the test data set) and reliably through the second wave of pandemic (AUC 0.84, 95% CI 0.84-0.85 on the prospective test data set). Given this data set was acquired later in time from September to November and more likely to suffer from data lag, the completeness and accuracy of outcome data are hypothesized to contribute to the decrease in model performance; a subgroup analysis on patients with the complete clinical features shows an improved performance (AUC 0.89, 95% CI 0.88-0.90; Table 3).

We also examined discriminatory capacity in subgroups stratified by sex, race, and age group separately. It predicts all-cause mortality similarly among men (AUC 0.84, 95% CI 0.84-0.84) and women (AUC 0.84, 95% CI 0.84-0.85) and is marginally more predictive among Asians (AUC 0.86, 95% CI 0.85-0.87) compared with African Americans (AUC 0.83, 95% CI 0.83-0.84) and Whites (AUC 0.84, 95% CI 0.84-0.84). Given age is an important predictor, the model is more sensitive toward elderly cohort (more accurately ruling out negative cases) and conversely more specific toward younger cohort (more accurately ruling in the positive cases).

Algorithm Selection

In the preliminary analysis, all the candidate algorithms perform comparably on test and prospective test data sets, with less than 3% difference in AUC for all outcomes between full and simplified models (Multimedia Appendix 4). Of the 6 candidate machine learning algorithms, boosting-based algorithms (XGB [10] and LightGBM [18]) performed consistently better [40] for both full and preliminary simplified models (n=20) with less computation time and produced well-calibrated probabilities (Multimedia Appendices 5 and 6); XGB was selected given it has been validated in a similar approach [7,41,42]. With adequate model calibration and low Brier score, no adjustment or calibration was subsequently performed.

Predictor Selection

Predictors were selected in the development pipeline; specifically, 100 individual runs of recursive predictor elimination are pooled at each step between 5- and 20-input model with an increment of 1. The selection of predictors was analyzed in the frequency heatmap (Multimedia Appendix 7) and automated from the pipeline while nonmodifiable factors such as diagnosis month or census division were precluded.

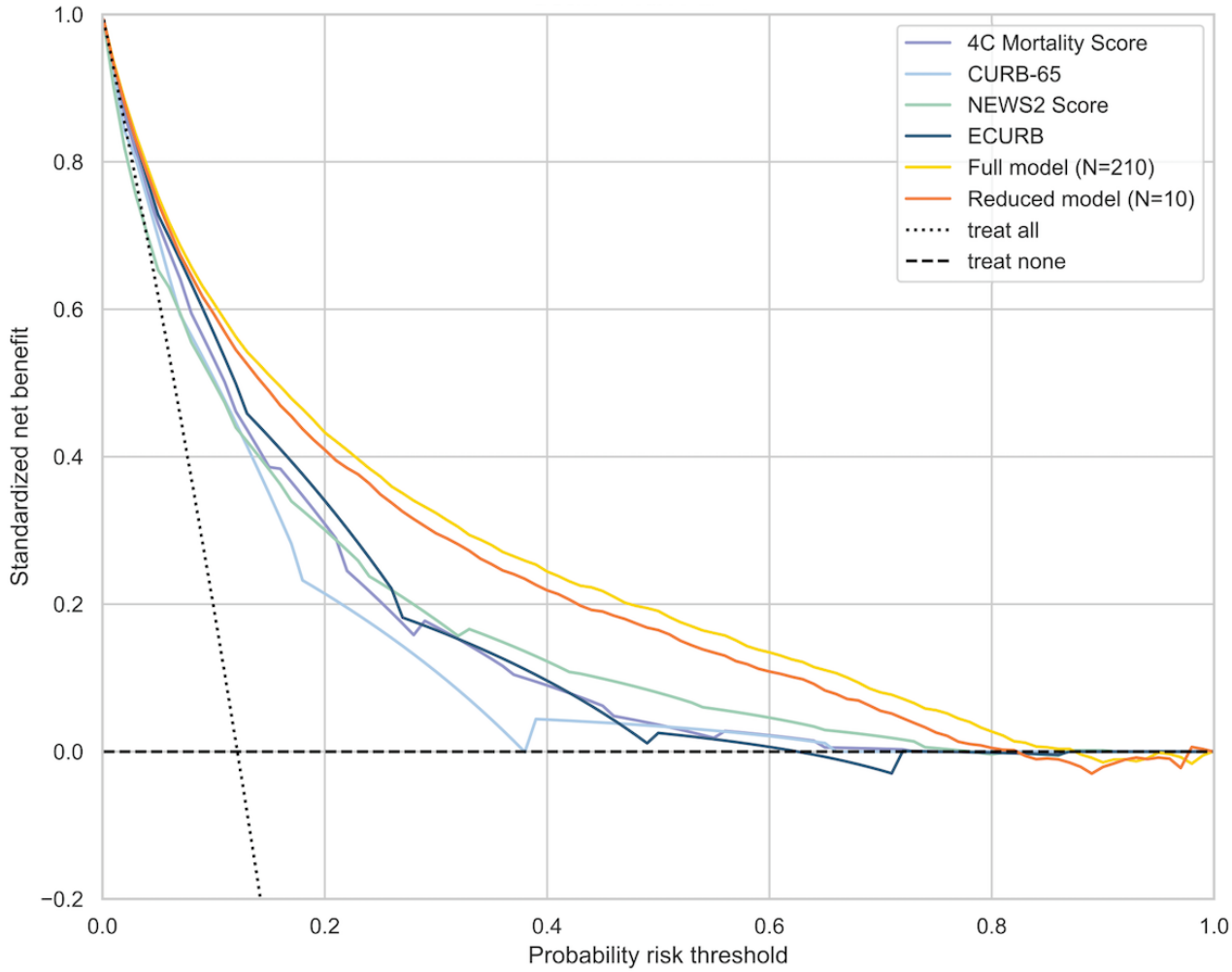
With only 10 predictors, the final parsimonious model (N10) still predicts COVID-19 adverse outcomes accurately and similarly to the full model (Table 2); for instance, the final parsimonious model consisting of age, systolic and diastolic blood pressures, respiration rate, pulse, temperature, BUN, SpO_2, albumin, and presence of any major cognitive disorder (including dementia, Parkinson disease, and Alzheimer disease) as input predicts all-cause mortality accurately with AUC of 0.88 (95% CI 0.87-0.88).

The magnitude and direction of individual feature contribution to prediction are inferred from the summary plot of Shapley values sorted by the descending order of feature impact (Multimedia Appendix 8); an increase in age [43-45], respiration rate [46], BUN [45,47,48], and aspartate transaminase [45,49], and a decrease in oxygen saturation [7,50], platelet count [26,51,52], and albumin [45,53] are associated with the increase in mortality risk.

Comparison With Existing Benchmark

The model shares commonalities (eg, age, respiration rate, blood pressures, pulse, BUN, SpO_2, albumin) with existing prognostic scores for community-acquired pneumonia or COVID-19 [5,26,27]; however, with automated feature selection from comprehensive input covariates, and machine learning algorithm, it compares favorably with existing scores across diagnostic statistics (Table 3) and shows greater clinical utility across a wide range of probability thresholds (Figure 5) in decision curve analysis.

Figure 5. Decision curve analysis of standardized net benefit across different risk thresholds. Dotted line represents the scenario if everyone is treated; dashed line represents the scenario if none is treated.



Discussion

Summary of Principal Findings

In this paper, we have adopted a systematic framework of developing and evaluating various machine learning techniques in predicting COVID-19 prognosis on near real-time, large-size EHR data in the United States. Boosting-based algorithms (XGBoost and LightGBM) have consistently outperformed other machine learning algorithms and COVID-19 benchmark risk scores with higher accuracy on the test data set (AUC 0.89, 95% CI 0.88-0.89) and on the prospective test data set (AUC 0.85, 95% CI 0.85-0.86), and better clinical utility on decision curve analysis. After further simplification of the model to only 10 clinical features, relative to full model it provides comparable discriminatory performance (AUC 0.88 95% CI 0.87-0.88) and clinical utility.

Predictors

A major strength of this study is the use of near real-time, large-size EHR data, resulting in predictors that are highly representative and relevant to clinical practice. We have restricted the analysis to commonly measured covariates with less than 30% (15,211/50,703) of missing values among the cohort. A higher coverage cutoff precludes key predictors such as oxygen saturation [7,50], respiration rate [46], and BUN [47,48] leading to degradation of model performance (Multimedia Appendices 9 and 10).

Postadmission treatment is not a major predictor of model performance; they are not included in the final analysis, which results in minimal impact of the model performance on all-cause mortality. Patients are presented at different disease trajectories when admitted to hospital, with some being in critical condition; for instance, among 9564 ICU patients, 4745 (49.61%) were admitted to ICU on day 1 of hospital admission. The relationship

between treatment type and outcome is therefore confounded by the stages of the disease course.

Age is identified as a crucial predictor for adverse outcomes [44]. It increases almost monotonically with health outcomes such as mortality and ARDS, but nonmonotonically with resource-dependent outcomes, such as ICU admission and invasive mechanical ventilation/ECMO, as these outcomes are closely associated with the availability of health care resources such as ventilator and ICU rooms. This is more noticeable for elderly patients over 75 years, who are disadvantaged for mechanical ventilation and ICU, presumably due to the scarcity of health care resources during the pandemic, though they are at highest mortality risk (Multimedia Appendix 11).

Clinical Application

When applying the model to clinical setting, threshold selection is of great practical importance in producing dichotomous predictions. In the data-driven, cost-agnostic approach, threshold is derived numerically from the AUC curve (Figure 3), which maximizes the Youden index [34] (P=.13). When the model is applied to inform clinical decision making, such as identifying patients for dexamethasone treatment, insights from relevant clinical trials could guide threshold calculation. For instance, the findings from the Randomised Evaluation of COVID-19 Therapy (RECOVERY) trial [54], a large-enrollment, randomized controlled trial of dexamethasone, indicate a mortality risk reduction of 4.84% among patients who received oxygen therapy or were mechanically ventilated (393/1603, 24.52%) compared with the control group (965/3287, 29.36%); conversely, there was an increase in mortality risk of 3.74% among patients who require no oxygen (dexamethasone group [89/501, 17.76%] vs usual care group [145/1034, 14.02%]). When the model is applied clinically as a prognostic tool to identify patients who will receive dexamethasone, cost of false negatives (ie, misclassifying patients as low risk, and therefore they missed dexamethasone treatment) is 4.84%, and cost of false positive (ie, misclassifying patients as high risk) is 3.74% when cost of misclassification is expressed as an increase in mortality risk. Given the mortality rate of 1592/6425 (24.78%) from the RECOVERY trial [54], the threshold is found from the AUC curve at P=.33, where the slope of curve [33,37] is (0.0374/0.0484) x (1 – 0.248)/(0.248) = 2.35. Model performances are evaluated at these 2 thresholds in Table 3.

Table 3. Comparison with existing risk scores evaluated on test data sets to predict 28-day all-cause mortality. Sensitivity and specificity were evaluated at 2 different thresholds.

Risk score	AUC[a] (95% CI), %	Threshold 1[b]		Threshold 2[c]		n[d]
		Sensitivity, %	Specificity, %	Sensitivity, %	Specificity, %	
Acute Physiology and Chronic Health Evaluation II	72.3 (69.5-74.9)	66.2	68.5	92.4	26.0	1769
Respiratory Rate-Oxygenation Index	68.5 (67.0-70.0)	28.2	92.7	54.2	78.3	16,640
CURB-65	78.7 (77.6-79.7)	36.2	92.4	77.2	69.1	15,001
E-CURB	81.9 (80.3-83.3)	63.4	83.4	87.3	61.3	5772
National Early Warning Score 2 score	82.9 (81.7-84.2)	51.6	91.2	75.0	77.0	14,112
Coronavirus Clinical Characterization Consortium Mortality score	82.2 (80.7-83.5)	62.3	83.8	71.8	75.7	6979
Baseline model	73.8 (73.2-74.5)	44.8	83.4	80.2	54.9	25,615
Full model	89.2 (88.1-90.3)	63.1	92.2	85.2	76.4	8493
N10 model	88.9 (88.0-90.0)	65.9	90.9	81.4	79.3	10,688

[a]AUC: area under the receiver operating characteristic curve.

[b]Threshold 1 is a clinically relevant threshold that identifies patients for dexamethasone treatment; costs of FP and FN are expressed in terms of mortality risk.

[c]Threshold 2 is derived from a cost-agnostic approach and is located at the point on the area under the receiver operating characteristic curve that maximizes the Youden index.

[d]Number of hospitalized patients in the test data set and the postdevelopment test data set with complete case.

These 2 thresholds (P=.13 and .24) are similar to the intermediate- and high-risk cutoffs used to define the severity of pneumonia [1,55,56]. Based on these approaches, we derived 2 clinically meaningful thresholds (Table 4), stratifying patients into (1) low-to-intermediate risk (P≤.13, observed mortality rate = 315/8065, 3.91%); (2) high risk (0.13<P≤.24, observed mortality rate = 225/1170, 19.23%); and (3) very high risk (P>.24, observed mortality rate = 787/1517, 51.88%). Scenario-based threshold can be substituted with appropriate clinical trial insights according to different treatment options.

Table 4. Mortality rate comparison across different risk groups on the test and postdevelopment prospective test data sets. Three risk groups were defined as (1) low-to-intermediate–risk group ($P \leq .13$), (2) high risk ($.13 < P \leq .24$), and (3) very high risk ($P > .24$). The threshold probabilities are obtained from receiver operating characteristic analysis, which (1) maximizes the Youden index ($P=.13$), or (2) defined by clinical utility of dexamethasone ($P=.24$) from the RECOVERY[a] trial.

Risk group	Test data set		Prospective test data set	
	Patients, n (%) (n=10,752)	Deaths, n (%) (n=1327)	Patients, n (%) (n=14,863)	Deaths, n (%) (n=1782)
Low–intermediate	8065 (75.01)	315 (3.91)	11,049 (74.34)	512 (4.63)
High	1170 (10.88)	225 (19.23)	1743 (11.73)	327 (18.76)
Very high	1517 (14.11)	787 (51.88)	2071 (13.93)	943 (45.53)

[a]RECOVERY: Randomised Evaluation of COVID-19 Therapy.

Strengths

The strengths of this research include the large size of data set, longitudinal nature, and near real-time update of the data release. The Optum database provides patient-level information with a diverse mix of geographic regions, insurance types, socioeconomic status, and ethnicity. A comprehensive list of 386 input covariates from baseline and at admission was included in the analysis based on epidemiological and clinical characteristics of COVID-19 cases; the end-to-end pipeline automates feature selection and model development process, producing risk factors that are both commonly measured at admission with wide coverage among study cohort and concordant with similar risk scores. This helps to improve the usability of the model without extensive electronic medical record integration or feeding the model with continuous data streams. The systematic approach and rigorous validations demonstrate consistent model performance to predict even beyond the period of data collection, with satisfactory discriminatory power and great clinical utility. Overall, the study offers an accurate, validated, and reliable prediction model based on only 10 clinical features as a prognostic tool for stratifying patients with COVID-19 into intermediate-, high-, and very high-risk groups. We envision this model to be used on the day of hospital admission at an inpatient setting where resource triaging is most relevant and early identification of high-risk patient is the key.

Limitations

There are several limitations in our study. First, the Optum COVID-19 database, being an EHR database, may not capture patients' entire interaction with health care systems because patients can switch between different hospitals or health care systems. This impacts several aspects of the study, from assessment of baseline comorbidity and comedication, to capture of outcomes during follow-up. Although we have identified a minimum of 10-week period from database refresh date to COVID-19 diagnosis date to allow for capture of follow-up data and outcomes, it is possible additional data lag is still present, challenging the completeness and accuracy of outcome assessment.

Because of Health Insurance Portability and Accountability Act (HIPAA)–compliance protection, patients over 89 years were included as a single category of age in the data set, with age being an important risk predictor of mortality. This can potentially lead to some performance degradation for patients aged over 89 years. Additional data, such as symptoms since onset, could aid in early prediction of aggressive COVID-19 progression, but these were available for less than 70% (35,493/50,703) of patients in the data set; if we had adequate data on patient symptoms and the use of oxygen therapies, model performance would likely improve. Similarly, this negatively impacts the evaluation of existing prognostic scores that require FiO_2. We have referred to the best currently available information on clinical trial for threshold calculation, and there could still exist differences in patient population between the RECOVERY trial and this work. Additional work is required for validating the results on vaccinated population.

Conclusions

In this study, we presented a systematic framework of model development based on a variety of machine learning techniques, combined with rigorous validation on statistically meaningful sample size. The model demonstrates consistent performance to predict even beyond the period of data collection. The parsimonious model with only 10 clinical features (age, systolic and diastolic blood pressures, respiration rate, pulse, temperature, BUN, SpO_2, albumin, and presence of major cognitive disorder) offers an accurate, validated, and calibrated prediction to stratifying COVID-19 patients into intermediate-, high-, and very high-risk groups. This simple predictive tool is shared with a wider health care community (Multimedia Appendix 12), to enable service as an early warning system to alert physicians of possible high-risk patients.

Acknowledgments

We acknowledge the extensive programming and planning work of the Amgen Center for Observational Research (Oana Abrahamian, Bagmeet Behera, Corinne Brooks, and Kimberly A Roehl), initial feasibility analysis and model hosting by Amgen Digital Health & Innovation - Data Science and Engineering team (Maxim Ivanov). We also acknowledge the health care

professionals whose tireless efforts in this unprecedented pandemic have provided critical knowledge, as well as the patients from whom we continue to learn so much.

Authors' Contributions

All the authors participated in literature search, conceptualization, data interpretation, reviewing, and editing the manuscript. JHP and FH were responsible for study design and methodology. FH performed the formal analysis and produced formatted tables and figures, and the original draft. KRW and AM conducted the initial feasibility analysis. All the authors have full access to the data in the study and accept responsibility to submit for publication.

Conflicts of Interest

FH, JHP, and AM are employees and stockholders of Amgen, Inc. KRW, an employee of League Inc, was formerly an employee of Amgen, Inc and owns stock in Amgen, Inc.

Multimedia Appendix 1
Model input and variable transformation. In the preliminary analysis, a total of 386 covariates with <30% missingness are incorporated as model input.
[PDF File (Adobe PDF File), 181 KB - jmir_v24i1e31549_app1.pdf]

Multimedia Appendix 2
Summary of missingness of vital and lab variables among the study cohort.
[PDF File (Adobe PDF File), 151 KB - jmir_v24i1e31549_app2.pdf]

Multimedia Appendix 3
Sensitivity analysis on model performances (AUC, 95% CI) between study cohort (n=50,703 in orange) and lab confirmed cohort (n=38,277 in blue) which is a subset of study cohort; (a) model performances on test dataset; (b) on post-development prospective test dataset.
[PDF File (Adobe PDF File), 615 KB - jmir_v24i1e31549_app3.pdf]

Multimedia Appendix 4
Model performance (AUC) during preliminary analysis (top: test dataset (N=10,752); bottom: post-development prospective test dataset (N=14,863). In the preliminary analysis, a total of 386 covariates (with <30% missingness) are incorporated as model input.
[PDF File (Adobe PDF File), 638 KB - jmir_v24i1e31549_app4.pdf]

Multimedia Appendix 5
Calibration curve during preliminary analysis (top panel: test dataset (N=10,752); bottom panel: post-development prospective test dataset (N=14,863)). In the preliminary analysis, a total of 386 covariates (with <30% missingness) are incorporated as model input.
[PDF File (Adobe PDF File), 591 KB - jmir_v24i1e31549_app5.pdf]

Multimedia Appendix 6
Runtime comparison among 6 candidate algorithms.
[PDF File (Adobe PDF File), 62 KB - jmir_v24i1e31549_app6.pdf]

Multimedia Appendix 7
Frequency heatmap of top 20 features at 5-variable to 20-variable algorithms (step size = 1). (a) all-cause mortality; (b) mechanical ventilation including ECMO; (c) ARDS including respiratory failure; (d) ICU admission during final analysis, excluding covariates of post-admission treatment.
[PDF File (Adobe PDF File), 670 KB - jmir_v24i1e31549_app7.pdf]

Multimedia Appendix 8
SHAP summary plot on 28-day mortality on aggregated datasets (test dataset and post-development prospective test dataset).
[PDF File (Adobe PDF File), 307 KB - jmir_v24i1e31549_app8.pdf]

Multimedia Appendix 9

Sensitivity analysis on model performances (AUC) between non-imputed (in orange) and imputed data (in blue); (a) 28-day mortality; (b) 28-day ICU admission; (c) composite of 28-day ARDS and respiratory failure; (d) composite of 28-day ECMO and invasive ventilation usage. Left – test dataset; right – post-development prospective test dataset.
[PDF File (Adobe PDF File), 175 KB - jmir_v24i1e31549_app9.pdf]

Multimedia Appendix 10
Sensitivity analysis on model performances (AUC) on non-imputed dataset with different thresholds of covariate coverage (10%, 30%, 50%, 70%, 80%, 90%) among the study cohort; (a) all-cause mortality; (b) ICU admission; (c) respiratory failure including ARDS; (d) invasive mechanical ventilation including ECMO. Left panel – test dataset; right panel – post-development prospective test dataset.
[PDF File (Adobe PDF File), 279 KB - jmir_v24i1e31549_app10.pdf]

Multimedia Appendix 11
SHAP dependence plot between age and four outcomes (top left: 28-day mortality; top right: composite of 28-day ARDS and respiratory failure; bottom left: 28-day ICU admission; bottom right: 28-day ECMO or ventilator. The features are colored by (a) minimum SpO_2 on admission; (b) respiration rate; (c) lymphocyte count; (d) BUN.
[PDF File (Adobe PDF File), 1546 KB - jmir_v24i1e31549_app11.pdf]

Multimedia Appendix 12
Gitlab repository.
[PDF File (Adobe PDF File), 49 KB - jmir_v24i1e31549_app12.pdf]

References

1. Knight SR, Ho A, Pius R, Buchan I, Carson G, Drake TM, ISARIC4C investigators. Risk stratification of patients admitted to hospital with covid-19 using the ISARIC WHO Clinical Characterisation Protocol: development and validation of the 4C Mortality Score. BMJ 2020 Sep 09;370:m3339 [FREE Full text] [doi: 10.1136/bmj.m3339] [Medline: 32907855]
2. Liang W, Liang H, Ou L, Chen B, Chen A, Li C, China Medical Treatment Expert Group for COVID-19. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness in Hospitalized Patients With COVID-19. JAMA Intern Med 2020 Aug 01;180(8):1081-1089 [FREE Full text] [doi: 10.1001/jamainternmed.2020.2033] [Medline: 32396163]
3. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985 Oct;13(10):818-829. [Medline: 3928249]
4. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonça A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996 Jul;22(7):707-710. [doi: 10.1007/BF01709751] [Medline: 8844239]
5. Chen J, Chang S, Liu JJ, Chan R, Wu J, Wang W, et al. Comparison of clinical characteristics and performance of pneumonia severity score and CURB-65 among younger adults, elderly and very old subjects. Thorax 2010 Nov 21;65(11):971-977. [doi: 10.1136/thx.2009.129627] [Medline: 20965934]
6. Zou X, Li S, Fang M, Hu M, Bian Y, Ling J, et al. Acute Physiology and Chronic Health Evaluation II Score as a Predictor of Hospital Mortality in Patients of Coronavirus Disease 2019. Crit Care Med 2020 Aug;48(8):e657-e665 [FREE Full text] [doi: 10.1097/CCM.0000000000004411] [Medline: 32697506]
7. Yadaw A, Li YC, Bose S, Iyengar R, Bunyavanich S, Pandey G. Clinical features of COVID-19 mortality: development and validation of a clinical prediction model. The Lancet Digital Health 2020 Oct;2(10):e516-e525 [FREE Full text] [doi: 10.1016/S2589-7500(20)30217-X] [Medline: 32984797]
8. Hajian-Tilaki K. Sample size estimation in diagnostic test studies of biomedical informatics. J Biomed Inform 2014 Apr;48:193-204 [FREE Full text] [doi: 10.1016/j.jbi.2014.02.013] [Medline: 24582925]
9. Buuren SV, Groothuis-Oudshoorn K. : Multivariate Imputation by Chained Equations in. J. Stat. Soft 2011;45(3):1-67. [doi: 10.18637/jss.v045.i03]
10. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. 2016 Aug 13 Presented at: KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA, USA p. 785-794.
11. Li WT, Ma J, Shende N, Castaneda G, Chakladar J, Tsai JC, et al. Using machine learning of clinical data to diagnose COVID-19: a systematic review and meta-analysis. BMC Med Inform Decis Mak 2020 Sep 29;20(1):247 [FREE Full text] [doi: 10.1186/s12911-020-01266-z] [Medline: 32993652]
12. Yu C, Lei Q, Li W, Wang X, Liu W, Fan X, et al. Clinical Characteristics, Associated Factors, and Predicting COVID-19 Mortality Risk: A Retrospective Study in Wuhan, China. Am J Prev Med 2020 Aug;59(2):168-175 [FREE Full text] [doi: 10.1016/j.amepre.2020.05.002] [Medline: 32564974]

13. Patrício A, Costa RS, Henriques R. Predictability of COVID-19 Hospitalizations, Intensive Care Unit Admissions, and Respiratory Assistance in Portugal: Longitudinal Cohort Study. J Med Internet Res 2021 Apr 28;23(4):e26075 [FREE Full text] [doi: 10.2196/26075] [Medline: 33835931]
14. Vaid A, Jaladanki SK, Xu J, Teng S, Kumar A, Lee S, et al. Federated Learning of Electronic Health Records to Improve Mortality Prediction in Hospitalized Patients With COVID-19: Machine Learning Approach. JMIR Med Inform 2021 Jan 27;9(1):e24207 [FREE Full text] [doi: 10.2196/24207] [Medline: 33400679]
15. Cheng F, Joshi H, Tandon P, Freeman R, Reich DL, Mazumdar M, et al. Using Machine Learning to Predict ICU Transfer in Hospitalized COVID-19 Patients. J Clin Med 2020 Jun 01;9(6):1668-1679 [FREE Full text] [doi: 10.3390/jcm9061668] [Medline: 32492874]
16. Parchure P, Joshi H, Dharmarajan K, Freeman R, Reich DL, Mazumdar M, et al. Development and validation of a machine learning-based prediction model for near-term in-hospital mortality among patients with COVID-19. BMJ Support Palliat Care 2020 Sep 22:1-8. [doi: 10.1136/bmjspcare-2020-002602] [Medline: 32963059]
17. Li Z, Wang L, Huang L, Zhang M, Cai X, Xu F, et al. Efficient management strategy of COVID-19 patients based on cluster analysis and clinical decision tree classification. Sci Rep 2021 May 05;11(1):9626 [FREE Full text] [doi: 10.1038/s41598-021-89187-3] [Medline: 33953307]
18. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. A highly efficient gradient boosting decision tree. In: Adv Neural Inf Process Syst 2017. LightGBM; 2017 Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4 - 7, 2017; Long Beach, CA, USA p. 3146-3154.
19. Guyon I, Weston J, Barnhill S, Vapnik V. Gene Selection for Cancer Classification using Support Vector Machines. Machine Learning 2002;46(1):389-422. [doi: 10.1023/A:1012487302797]
20. Brier GW. Verification of Forecasts Expressed in Terms of Probability. Monthly Weather Review 1950 Jan;78(1):1-3. [doi: 10.1175/1520-0493(1950)078<0001:vofeit>2.0.co;2]
21. Niculescu-Mizil A, Caruana R. Predicting Good Probabilities with Supervised Learning. In: Proceedings of the 22 nd International Conference on Machine Learning. USA: Association for Computing Machinery; 2005 Presented at: Proceedings of the 22nd International Conference on Machine Learning; Aug 7 - 11, 2005; New York, NY p. 625-632. [doi: 10.1145/1102351.1102430]
22. Goodhue, Lewis, Thompson. Does PLS Have Advantages for Small Sample Size or Non-Normal Data? MIS Quarterly 2012;36(3):981. [doi: 10.2307/41703490]
23. Stephens JR, Stümpfle R, Patel P, Brett S, Broomhead R, Baharlo B, et al. Analysis of Critical Care Severity of Illness Scoring Systems in Patients With Coronavirus Disease 2019: A Retrospective Analysis of Three U.K. ICUs. Crit Care Med 2021 Jan 01;49(1):e105-e107 [FREE Full text] [doi: 10.1097/CCM.0000000000004674] [Medline: 32991357]
24. Capelastegui A, España PP, Quintana JM, Areitio I, Gorordo I, Egurrola M, et al. Validation of a predictive rule for the management of community-acquired pneumonia. Eur Respir J 2006 Jan;27(1):151-157 [FREE Full text] [doi: 10.1183/09031936.06.00062505] [Medline: 16387948]
25. Wellbelove Z, Walsh C, Perinpanathan T, Lillie P, Barlow G. Comparing the 4C mortality score for COVID-19 to established scores (CURB65, CRB65, qSOFA, NEWS) for respiratory infection patients. J Infect 2021 Mar;82(3):414-451 [FREE Full text] [doi: 10.1016/j.jinf.2020.10.015] [Medline: 33115655]
26. Liu J, Xu F, Hui Zhou, Wu X, Shi L, Lu R, et al. Expanded CURB-65: a new score system predicts severity of community-acquired pneumonia with superior efficiency. Sci Rep 2016 Mar 18;6(1):22911 [FREE Full text] [doi: 10.1038/srep22911] [Medline: 26987602]
27. National Early Warning Score (NEWS) 2. URL: https://www.rcplondon.ac.uk/projects/outputs/national-early-warning-score-news-2 [accessed 2020-11-17]
28. Carr E, Bendayan R, Bean D, O´Gallagher K, Pickles A, Stahl D, et al. Evaluation and improvement of the National Early Warning Score (NEWS2) for COVID-19: a multi-hospital study. BMC Med 2021 Jan 21;19(1):23. [doi: 10.1186/s12916-020-01893-3] [Medline: 33472631]
29. Prower E, Grant D, Bisquera A, Breen CP, Camporota L, Gavrilovski M, et al. The ROX index has greater predictive validity than NEWS2 for deterioration in Covid-19. EClinicalMedicine 2021 May;35:100828 [FREE Full text] [doi: 10.1016/j.eclinm.2021.100828] [Medline: 33937729]
30. Roca O, Messika J, Caralt B, García-de-Acilu M, Sztrymf B, Ricard J, et al. Predicting success of high-flow nasal cannula in pneumonia patients with hypoxemic respiratory failure: The utility of the ROX index. J Crit Care 2016 Oct;35:200-205 [FREE Full text] [doi: 10.1016/j.jcrc.2016.05.022] [Medline: 27481760]
31. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From Local Explanations to Global Understanding with Explainable AI for Trees. Nat Mach Intell 2020 Jan;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
32. Lundberg S, Lee S. A Unified Approach to Interpreting Model Predictions. 2017 Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); Dec 4 - 9, 2017; Long Beach, CA, USA URL: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

33. Habibzadeh F, Habibzadeh P, Yadollahie M. On determining the most appropriate test cut-off value: the case of tests with continuous results. Biochem Med (Zagreb) 2016 Oct 15;26(3):297-307 [FREE Full text] [doi: 10.11613/BM.2016.034] [Medline: 27812299]
34. Youden W. Index for rating diagnostic tests. Cancer 1950 Jan;3(1):32-35. [doi: 10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.co;2-3] [Medline: 15405679]
35. Schisterman EF, Perkins NJ, Liu A, Bondell H. Optimal cut-point and its corresponding Youden Index to discriminate individuals using pooled blood samples. Epidemiology 2005 Jan;16(1):73-81. [doi: 10.1097/01.ede.0000147512.81966.ba] [Medline: 15613948]
36. Shapiro DE. The interpretation of diagnostic tests. Stat Methods Med Res 1999 Jun;8(2):113-134. [doi: 10.1177/096228029900800203] [Medline: 10501649]
37. Greiner M, Pfeiffer D, Smith R. Principles and practical application of the receiver-operating characteristic analysis for diagnostic tests. Preventive Veterinary Medicine 2000 May;45(1-2):23-41 [FREE Full text] [doi: 10.1016/s0167-5877(00)00115-x]
38. Vickers AJ, Van Calster B, Steyerberg EW. Net benefit approaches to the evaluation of prediction models, molecular markers, and diagnostic tests. BMJ 2016 Jan 25;352:i6 [FREE Full text] [doi: 10.1136/bmj.i6] [Medline: 26810254]
39. Vickers AJ, van Calster B, Steyerberg EW. A simple, step-by-step guide to interpreting decision curve analysis. Diagn Progn Res 2019;3:18 [FREE Full text] [doi: 10.1186/s41512-019-0064-7] [Medline: 31592444]
40. Subudhi S, Verma A, Patel AB, Hardin CC, Khandekar MJ, Lee H, et al. Comparing machine learning algorithms for predicting ICU admission and mortality in COVID-19. NPJ Digit Med 2021 May 21;4(1):87 [FREE Full text] [doi: 10.1038/s41746-021-00456-x] [Medline: 34021235]
41. Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, Northwell COVID-19 Research Consortium. A Machine Learning Prediction Model of Respiratory Failure Within 48 Hours of Patient Admission for COVID-19: Model Development and Validation. J Med Internet Res 2021 Feb 10;23(2):e24246 [FREE Full text] [doi: 10.2196/24246] [Medline: 33476281]
42. Kim H, Han D, Kim J, Kim D, Ha B, Seog W, et al. An Easy-to-Use Machine Learning Model to Predict the Prognosis of Patients With COVID-19: Retrospective Cohort Study. J Med Internet Res 2020 Nov 09;22(11):e24225 [FREE Full text] [doi: 10.2196/24225] [Medline: 33108316]
43. Challen R, Brooks-Pollock E, Read JM, Dyson L, Tsaneva-Atanasova K, Danon L. Risk of mortality in patients infected with SARS-CoV-2 variant of concern 202012/1: matched cohort study. BMJ 2021 Mar 09;372:n579 [FREE Full text] [doi: 10.1136/bmj.n579] [Medline: 33687922]
44. O'Driscoll M, Ribeiro Dos Santos G, Wang L, Cummings DAT, Azman AS, Paireau J, et al. Age-specific mortality and immunity patterns of SARS-CoV-2. Nature 2021 Feb;590(7844):140-145. [doi: 10.1038/s41586-020-2918-0] [Medline: 33137809]
45. Zhou J, Lee S, Wang X, Li Y, Wu WKK, Liu T, et al. Development of a multivariable prediction model for severe COVID-19 disease: a population-based study from Hong Kong. NPJ Digit Med 2021 Apr 08;4(1):66 [FREE Full text] [doi: 10.1038/s41746-021-00433-4] [Medline: 33833388]
46. Natarajan A, Su H, Heneghan C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. NPJ Digit Med 2020 Nov 30;3(1):156 [FREE Full text] [doi: 10.1038/s41746-020-00363-7] [Medline: 33299095]
47. Ok F, Erdogan O, Durmus E, Carkci S, Canik A. Predictive values of blood urea nitrogen/creatinine ratio and other routine blood parameters on disease severity and survival of COVID-19 patients. J Med Virol 2021 Feb;93(2):786-793 [FREE Full text] [doi: 10.1002/jmv.26300] [Medline: 32662893]
48. Cheng A, Hu L, Wang Y, Huang L, Zhao L, Zhang C, et al. Diagnostic performance of initial blood urea nitrogen combined with D-dimer levels for predicting in-hospital mortality in COVID-19 patients. Int J Antimicrob Agents 2020 Sep;56(3):106110 [FREE Full text] [doi: 10.1016/j.ijantimicag.2020.106110] [Medline: 32712332]
49. Marjot T, Webb GJ, Barritt AS, Moon AM, Stamataki Z, Wong VW, et al. COVID-19 and liver disease: mechanistic and clinical perspectives. Nat Rev Gastroenterol Hepatol 2021 May;18(5):348-364 [FREE Full text] [doi: 10.1038/s41575-021-00426-4] [Medline: 33692570]
50. Razavian N, Major VJ, Sudarshan M, Burk-Rafel J, Stella P, Randhawa H, et al. A validated, real-time prediction model for favorable outcomes in hospitalized COVID-19 patients. NPJ Digit Med 2020;3:130 [FREE Full text] [doi: 10.1038/s41746-020-00343-x] [Medline: 33083565]
51. Liu H, Chen J, Yang Q, Lei F, Zhang C, Qin J, et al. Development and validation of a risk score using complete blood count to predict in-hospital mortality in COVID-19 patients. Med (NY) 2021 Apr 09;2(4):435-447.e4 [FREE Full text] [doi: 10.1016/j.medj.2020.12.013] [Medline: 33521746]
52. Bi X, Su Z, Yan H, Du J, Wang J, Chen L, et al. Prediction of severe illness due to COVID-19 based on an analysis of initial Fibrinogen to Albumin Ratio and Platelet count. Platelets 2020 Jul 03;31(5):674-679 [FREE Full text] [doi: 10.1080/09537104.2020.1760230] [Medline: 32367765]
53. Kheir M, Saleem F, Wang C, Mann A, Chua J. Higher albumin levels on admission predict better prognosis in patients with confirmed COVID-19. PLoS One 2021;16(3):e0248358 [FREE Full text] [doi: 10.1371/journal.pone.0248358] [Medline: 33725003]

54. RECOVERY Collaborative Group, Horby P, Lim WS, Emberson JR, Mafham M, Bell JL, et al. Dexamethasone in Hospitalized Patients with Covid-19. N Engl J Med 2021 Feb 25;384(8):693-704 [FREE Full text] [doi: 10.1056/NEJMoa2021436] [Medline: 32678530]
55. Lim WS, van der Eerden MM, Laing R, Boersma WG, Karalus N, Town GI, et al. Defining community acquired pneumonia severity on presentation to hospital: an international derivation and validation study. Thorax 2003 May;58(5):377-382 [FREE Full text] [doi: 10.1136/thorax.58.5.377] [Medline: 12728155]
56. Kaysin A, Viera AJ. Community-Acquired Pneumonia in Adults: Diagnosis and Management. Am Fam Physician 2016 Nov 01;94(9):698-706 [FREE Full text] [Medline: 27929242]

Edited by G Eysenbach; submitted 24.06.21; peer-reviewed by HJ Kim; comments to author 16.07.21; revised version received 26.07.21; accepted 19.12.21; published 21.01.22.

Please cite as:
He F, Page JH, Weinberg KR, Mishra A
The Development and Validation of Simplified Machine Learning Algorithms to Predict Prognosis of Hospitalized Patients With COVID-19: Multicenter, Retrospective Study
J Med Internet Res 2022;24(1):e31549
URL: https://www.jmir.org/2022/1/e31549
doi:10.2196/31549
PMID:34951865

Original Paper

2B-Alert Web 2.0, an Open-Access Tool for Predicting Alertness and Optimizing the Benefits of Caffeine: Utility Study

Jaques Reifman[1], PhD; Kamal Kumar[1,2], MS; Luke Hartman[1,2], BS; Andrew Frock[1,2], BS; Tracy J Doty[3], PhD; Thomas J Balkin[3,4], PhD; Sridhar Ramakrishnan[1,2], PhD; Francisco G Vital-Lopez[1,2], PhD

[1]Department of Defense Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Development Command, Fort Detrick, MD, United States

[2]The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc, Bethesda, MD, United States

[3]Behavioral Biology Branch, Walter Reed Army Institute of Research, Silver Spring, MD, United States

[4]Oak Ridge Institute for Science and Education, Research Participation Program, Oak Ridge, TN, United States

Corresponding Author:
Jaques Reifman, PhD
Department of Defense Biotechnology High Performance Computing Software Applications Institute
Telemedicine and Advanced Technology Research Center
U.S. Army Medical Research and Development Command
504 Scott Street
Fort Detrick, MD, 21702-5012
United States
Phone: 1 301 619 7915
Fax: 1 301 619 1983
Email: jaques.reifman.civ@mail.mil

Abstract

Background: One-third of the US population experiences sleep loss, with the potential to impair physical and cognitive performance, reduce productivity, and imperil safety during work and daily activities. Computer-based fatigue-management systems with the ability to predict the effects of sleep schedules on alertness and identify safe and effective caffeine interventions that maximize its stimulating benefits could help mitigate cognitive impairment due to limited sleep. To provide these capabilities to broad communities, we previously released *2B-Alert* Web, a publicly available tool for predicting the average alertness level of a group of individuals as a function of time of day, sleep history, and caffeine consumption.

Objective: In this study, we aim to enhance the capability of the *2B-Alert* Web tool by providing the means for it to *automatically* recommend safe and effective caffeine interventions (time and dose) that lead to optimal alertness levels at user-specified times under any sleep-loss condition.

Methods: We incorporated a recently developed caffeine-optimization algorithm into the predictive models of the original *2B-Alert* Web tool, allowing the system to search for and identify viable caffeine interventions that result in user-specified alertness levels at desired times of the day. To assess the potential benefits of this new capability, we simulated four sleep-deprivation conditions (sustained operations, restricted sleep with morning or evening shift, and night shift with daytime sleep) and compared the alertness levels resulting from the algorithm's recommendations with those based on the US Army caffeine-countermeasure guidelines. In addition, we enhanced the usability of the tool by adopting a drag-and-drop graphical interface for the creation of sleep and caffeine schedules.

Results: For the 4 simulated conditions, the *2B-Alert* Web–proposed interventions increased mean alertness by 36% to 94% and decreased peak alertness impairment by 31% to 71% while using equivalent or smaller doses of caffeine as the corresponding US Army guidelines.

Conclusions: The enhanced capability of this evidence-based, publicly available tool increases the efficiency by which diverse communities of users can identify safe and effective caffeine interventions to mitigate the effects of sleep loss in the design of research studies and work and rest schedules.

(J Med Internet Res 2022;24(1):e29595) doi:10.2196/29595

KEYWORDS

alertness-prediction model; caffeine intervention; neurobehavioral performance; psychomotor vigilance test; PVT; sleep loss

Introduction

Background

Previously, we developed and publicly released the *2B-Alert* Web application [1], allowing users to compare and contrast predictions of alertness levels based on the psychomotor vigilance test (PVT) for a group of individuals as a function of time of day, sleep and wake schedule, and caffeine dose [2]. Over the last 15 years, our group at the US Army incrementally developed and enhanced mathematical models that form the core of the *2B-Alert* Web tool. At each developmental step, we created models with additional capabilities and independently validated the model predictions using an array of studies that investigated different sleep-deprivation conditions [3-8]. In total, we have assessed the model predictions for nearly 1200 participants from more than 50 laboratory and field studies. The study conditions ranged from sleep extension (10 h [hours] in bed) to chronic sleep restriction (3-7 h of sleep per night) to total sleep deprivation (TSD; up to 88 h of continuous wakefulness), with several studies performed to investigate the recuperative effects of caffeine administered in both single and multiple doses (50-600 mg), under a variety of sleep and wake schedules.

To date, the *2B-Alert* Web tool has nearly 25,000 registered users from 144 countries, with more than 1800 users from 54 countries accessing the site ≥2 times in the last 12 months and a daily average of 26 log-ins. This is the only publicly available tool of its kind. Here, we describe an enhanced version of the tool, which has the added capability of automatically suggesting caffeine interventions (time and dose) to optimize alertness levels for desired (user-specified) times of day. For any particular combination of user-defined sleep and wake schedule, desired peak-alertness periods, maximum alertness-impairment threshold during these periods, and maximum total caffeine consumption in a 24-h period, the updated tool generates an optimal and safe caffeine-dosing schedule. Specifically, the algorithm generates schedules that use the least amount of caffeine to achieve a user-defined alertness level or achieve *maximum* alertness levels for a user-defined amount of caffeine [9]. This capability would be particularly important during sleep-deprivation conditions because it maximizes the utility of caffeine as a fatigue countermeasure.

Caffeine-Optimization Algorithm

To attain this functionality, we incorporated a recently developed caffeine-optimization algorithm [9] with the predictive models [7,8] of the original *2B-Alert* Web application [2]. Previously, the tool predicted alertness levels for user-defined sleep and wake and caffeine schedules but required multiple trial-and-error simulations when the user wanted to determine a caffeine schedule that resulted in peak alertness levels during desired wake periods. In the new, enhanced version, this process is performed automatically, leading to a more effective means to identify safe caffeine interventions to guide the design of work and rest schedules and caffeine studies. Using this new capability of the tool, we demonstrated that, compared with the US Army guidelines for the use of caffeine as a countermeasure to sleep deprivation [10], the *2B-Alert* Web–proposed interventions increased mean alertness by 36% to 94% and decreased peak impairment by 31% to 71% while using equivalent or smaller doses of caffeine. In addition, we enhanced the usability of the tool by adopting a drag-and-drop graphical interface for the creation of sleep and caffeine schedules.

Methods

Original and Updated Capabilities

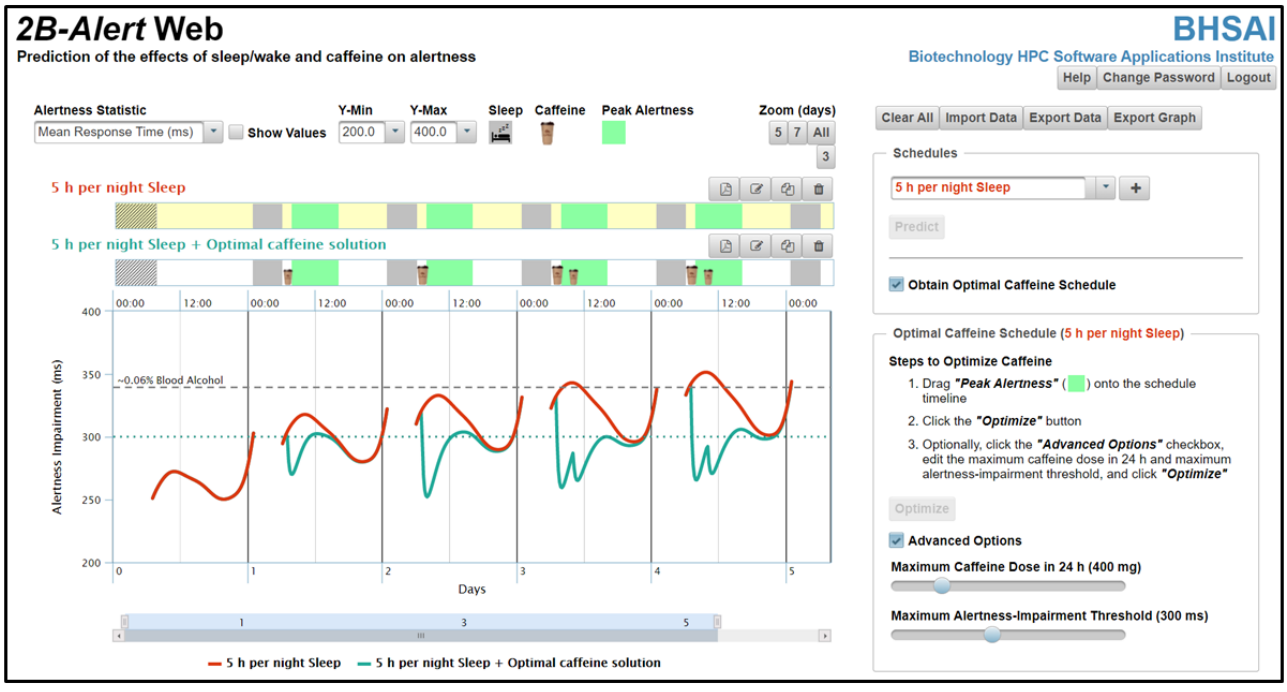
The enhanced *2B-Alert* Web tool predicts PVT performance, a measure of alertness and sustained attention, for a typical individual (as determined from group data) as a function of time of day, sleep and wake schedule, and caffeine consumption (dose and time). It offers the capability to accomplish the following:

- Compare and contrast the effects of different sleep and wake and caffeine schedules on alertness.
- Automatically identify caffeine interventions that lead to the desired alertness levels at the desired times of day.

For the former capability, which is also offered in the original tool [2], it provides alertness-level (ie, PVT) predictions for a given sleep and wake and caffeine schedule (Figure 1). Now, the updated tool offers the optional capability to automatically identify the optimal time and dose of caffeine consumption that lead to the desired, user-defined alertness outcomes. For (1) a specified sleep and wake schedule, (2) desired periods of peak alertness, (3) maximum acceptable alertness-impairment threshold during peak-alertness periods, and (4) maximum total caffeine consumption over a 24-h running period, the tool provides caffeine timing and dosing suggestions to achieve peak alertness at the desired times to the extent allowed by the limit on caffeine consumption (Figure 2).

Figure 1. Log-in screen of the *2B-Alert* Web application. This initial screen is preloaded with 3 sleep and caffeine schedules. The yellow background indicates the selected schedule (5 h per night sleep). Users select a schedule by clicking on it and add sleep and caffeine episodes by dragging and dropping the corresponding icon at the top of the screen onto the schedule timeline. The colors of the alertness-impairment prediction plots match those of the names of the corresponding schedules, and users select through a drop-down menu one of three predicted alertness outcome statistics for the psychomotor vigilance test: mean response time (shown), mean speed, or number of lapses >500 ms for a 10-minute psychomotor vigilance test. Users can obtain brief descriptions of the graphical interface functionalities by hovering over the various buttons with the computer mouse. A more comprehensive description is available in the user guide accessible by pressing the Help button at the upper right-hand corner of the page. h: hours.

Figure 2. Caffeine-optimization screen. Clicking on the *Obtain Optimal Caffeine Schedule* checkbox at the bottom of the right-hand panel in Figure 1 takes users to the caffeine-optimization screen shown here, and unchecking the box returns users to the prediction screen in Figure 1. From this screen, users can define periods of peak alertness by dragging and dropping the green icon at the top of the page onto the selected schedule and obtain caffeine schedules that result in desired alertness levels for the user-specified peak-alertness periods. The Advanced Options checkbox (lower right-hand panel) allows users to specify thresholds in the optimization algorithm for total caffeine in a running 24-h period and alertness-impairment levels during periods of peak alertness. h: hours.



Overview of the 2B-Alert Web Tool

Upon user log-in, the tool provides brief stepwise instructions on how to select, edit, and add sleep and caffeine episodes to a schedule and make predictions (Figure 1, right-hand panel). Similarly, when the user navigates to the caffeine optimization screen by clicking the *Obtain Optimal Caffeine Schedule* checkbox at the bottom of the right-hand panel in Figure 1, the tool also provides brief stepwise instructions on how to create peak alertness periods, obtain an optimal caffeine schedule, and how to edit the alertness-impairment threshold and the maximum total caffeine consumption over a 24-h running period (Figure 2, right-hand panel). In addition, users can obtain brief descriptions of the functionalities of the various interaction-enabled buttons of the graphical interface by hovering over the buttons with a mouse. When the user clicks on the *Help* button in the upper right-hand corner of the display window, the system displays a user guide, which provides a comprehensive description of the functionalities of the system to facilitate assimilation of the *2B-Alert* Web tool.

Comparing and Contrasting Different Sleep and Caffeine Schedules

The *2B-Alert* Web tool can be used to compare and contrast the effects of different schedules on alertness. A schedule is defined by a series of sleep episodes and caffeine episodes spanning a period of up to 30 days (note that all day and time entries are for the same time zone). In turn, a sleep episode is defined by start and end days and times, whereas a caffeine episode is defined by the day, time, and dose of caffeine, with the dose entered manually or selected from a drop-down list of more than 30 popular caffeine-containing products. Upon user log-in, the tool displays three preloaded schedules: *8 h per night sleep*, *5 h per night sleep*, and *5 h per night sleep + 100 mg caffeine*, with the middle schedule (yellow background) selected, as shown in the screenshot of the *2B-Alert* Web tool interface (Figure 1).

The user can modify an existing schedule or create a new schedule. For an existing schedule, users may view and edit the day and time of sleep episodes (gray icons) as well as caffeine-dose episodes (cup icons) by clicking on the corresponding icon. For example, when the user clicks on the sleep icon of the selected schedule in Figure 1 (*5 h per night sleep*), a pop-up window displays the start and end days of the sleep episode, along with the start (01:00) and end (06:00) times. Users may also add sleep and caffeine episodes by dragging and dropping the corresponding icon located above the schedules onto the timeline of the selected schedule. Users may select a schedule by clicking on its timeline or by choosing it from the drop-down list of preloaded timelines under *Schedules* on the right-hand side of the figure. The four buttons, from left to right, on the upper right-hand side of each schedule timeline allow users to export the schedule as a PDF file, change the schedule name, save a schedule with the same or a different name, or delete the schedule, respectively.

To create a new schedule, the user clicks the plus button to the right of the Schedule drop-down menu (Figure 1, right-hand panel) and then adds sleep and caffeine episodes. Alternatively, a schedule can be imported from a Microsoft Excel file (Figure 1, *Import Data*) using a predefined format. Each generated or edited schedule can be saved with a user-defined name and exported as a PDF file (Figure 1, left-most of the 4 buttons on the upper right-hand side of this schedule's timeline). The system supports up to 5 schedules and plots per session, where users can hide (or unhide) plots from view by clicking on the corresponding schedule name below the x-axis scroll bar. If the interface is loaded with 5 schedules, to create a new schedule, the user should delete an existing one.

The predicted alertness impairment for each schedule can be compared in the graph below the schedules. Figure 1 shows the 3 corresponding predictions for the 3 preloaded schedules using the mean response time (RT; in ms) PVT statistic, from *day 0* through *day 4* of the 4-day schedule, which starts on *day 1* and lasts until the start of the sleep episode on *day 5*. The plots allow for the comparison of the effects of different sleep durations on alertness (eg, 5 h vs 8 h of sleep per night), as well as for the assessment of the beneficial effects of caffeine countermeasures (eg, 5 h per night of sleep with and without 100 mg of daily caffeine at 08:00). The displayed plots can be saved in an image file using the *Export Graph* button, and the numerical values for each of the 3 predicted statistics can be exported into an Excel file, along with the corresponding sleep and wake and caffeine schedules (Figure 1, *Export Data*). This allows users to import and reuse schedules in a future session because the system does not save schedules or their predictions on the web (the system erases all data when the user logs out).

In addition to the mean RT, the user can select to display the plots of alertness-impairment predictions for two other PVT statistics: mean speed (average of the reciprocal of RT; in 1/s) and number of lapses (number of RTs >500 ms). The statistics are for a 10-minute test and are selected from the *Alertness Statistic* drop-down menu located above the schedules (Figure 1). To map PVT statistics into a more broadly understood metric of vigilance deficits, we used the findings from Dawson and Reid [11] and Williamson et al [12] to obtain an equivalence between PVT alertness-impairment values and blood alcohol concentrations (BACs). We estimated that a mean RT of 339 ms attained after 19 h of continued wakefulness corresponded to a 0.06% BAC (Figure 1, horizontal dashed gray line) and that a mean RT of 458 ms attained after 24 h of continuous wakefulness corresponded to a 0.08% BAC. For context, driving at BACs of 0.06% and 0.08% increase the risk of causing a traffic accident by 2- and 3-fold, respectively, compared with control drivers [13,14].

Automatic Identification of Caffeine Interventions

To automatically identify caffeine interventions that lead to desired alertness levels for a selected sleep and wake schedule, the user should click the *Obtain Optimal Caffeine Schedule* checkbox at the bottom of the right-hand panel in Figure 1, which takes users to another graphical interface (Figure 2). In this interface, the user can add a desired period of peak alertness (start and end days and times), which is a required input, to the schedule. This is achieved by dragging and dropping the *Peak Alertness* icon (green icon) located above the schedules onto the timeline (Figure 2). Next, clicking the *Optimize* button in the lower right-hand panel will generate the optimal caffeine

schedule and display the corresponding alertness-impairment prediction. For the selected schedule shown in Figure 1 (*5 h per night sleep*), these steps resulted in the identification of the optimal caffeine intervention named *5 h per night Sleep + Optimal caffeine solution* in Figure 2 and the corresponding alertness-impairment prediction plot. In this case, peak-alertness periods from 08:00 to 16:00 for *days 1* to *4* required a single caffeine dose at 07:00 of 100 mg on *day 1* and 200 mg on *day 2* and two doses at 07:00 and 10:00 of 200 mg and 100 mg, respectively, on *days 3* and *4* (see Figure S1 in Multimedia Appendix 1 for details; larger cup icons indicate larger caffeine doses). Optionally, the *Advanced Options* checkbox at the bottom of the lower right-hand panel allows users to set thresholds for maximum total caffeine in a 24-h period (100-1500 mg; default 400 mg) and the maximum alertness-impairment level (RT ranging from 150-500 ms; default 300 ms) by dragging the sliders to the desired values (Figure 2). As described previously, users can rename, save, delete, and export the updated schedule, as well as save the displayed plot as an image file. Unchecking the *Obtain Optimal Caffeine Schedule* in the upper right-hand panel takes users back to the prediction graphical interface (Figure 1).

Initial Conditions and Assumptions

We formulated the predictive model in the *2B-Alert* Web tool so that alertness is an inversely related function of sleep debt, which accumulates over days with <8 h of sleep per day [5,7,8] but decreases for sleep durations of >8 h per day. We initialized the tool so that on day *0* there is no sleep debt after 8 h of sleep (23:00-07:00). However, if sleep debt is nonzero, users need to enter up to 7 days of sleep history at the beginning of the schedule (sleep episodes >7 days old have negligible influence on near-future alertness [5,7,8]).

We assumed that the restorative effect of caffeine depends on the alertness level; that is, for a given caffeine dose, the larger the alertness impairment, the greater the beneficial effect of caffeine [15,16], where the magnitude of the benefit depends on the impairment level, the caffeine dose, and the residual concentration of caffeine from previous doses.

For the automatic identification of caffeine interventions that optimize alertness levels, we formulated the optimization problem so that, for the desired periods of peak alertness, the algorithm seeks caffeine-intervention solutions that equally weigh the cumulative deficit above the maximum alertness-impairment threshold and the peak alertness-impairment level [9]. To obtain practical and safe solutions, we added the following constraints: (1) caffeine doses to be restricted to 100 mg, 200 mg, or 300 mg; (2) dosing to occur on the hour; (3) the minimum time period between doses to be 2 h; and (4) the cumulative caffeine concentration in the blood to be below the maximum level achieved by a single 400-mg dose [17]. The optimization algorithm finds solutions in a matter of seconds by generating and assessing only caffeine schedules that are likely to reduce alertness impairment [9]. The tool attempts to find solutions that use the least amount of caffeine, while meeting the imposed constraints. However, in certain cases, even when using the maximum amount of caffeine, it may not be possible to obtain solutions that reduce the alertness impairment below the maximum impairment threshold. To obtain solutions with lower alertness impairment, users may increase the maximum total caffeine consumption in a 24-h period (Figure 2, *Advanced Options*). Alternatively, the near-optimal *2B-Alert* Web solution provides a good starting point for manual exploration of additional caffeine interventions that are potentially more effective but likely violate specified constraints.

Simulations to Assess the Optimization Algorithm

To assess the efficacy and potential benefits of the automated caffeine interventions proposed by the *2B-Alert* Web algorithm, we performed 4 simulations encompassing the three circumstances under which caffeine countermeasures are considered by the US Army guidelines [10]: sustained operations, restricted sleep, and night shift. Table 1 summarizes the US Army guidelines for the use of caffeine as a countermeasure to sleep deprivation and the simulated conditions, including the sleep-deprivation schedules and the assumed periods of peak alertness. For example, for the sustained operations scenario in condition 1, we simulated a 30-h TSD challenge from 07:00 on *day1* to 13:00 on *day2* and arbitrarily assumed a 13-h peak-alertness period from 00:00 to 13:00 on *day2* and a maximum alertness-impairment threshold of 270 ms (Figure 3), which corresponds to the maximum alertness impairment under well-rested conditions (ie, habitual sleep of 8 h per day). Then, we used the tool to obtain an optimal caffeine schedule and compared the resulting alertness impairment with that obtained by using the US Army guidelines. Figures S2-S7 in Multimedia Appendix 1 provide details for the simulations of conditions 2-4.

Table 1. Summary of the US Army guidelines for the use of caffeine as a countermeasure to sleep deprivation, simulated conditions (including type of sleep challenge, sleep schedule, and desired period of peak alertness), and recommended caffeine countermeasures based on the guidelines and the *2B-Alert* Web optimization algorithm.

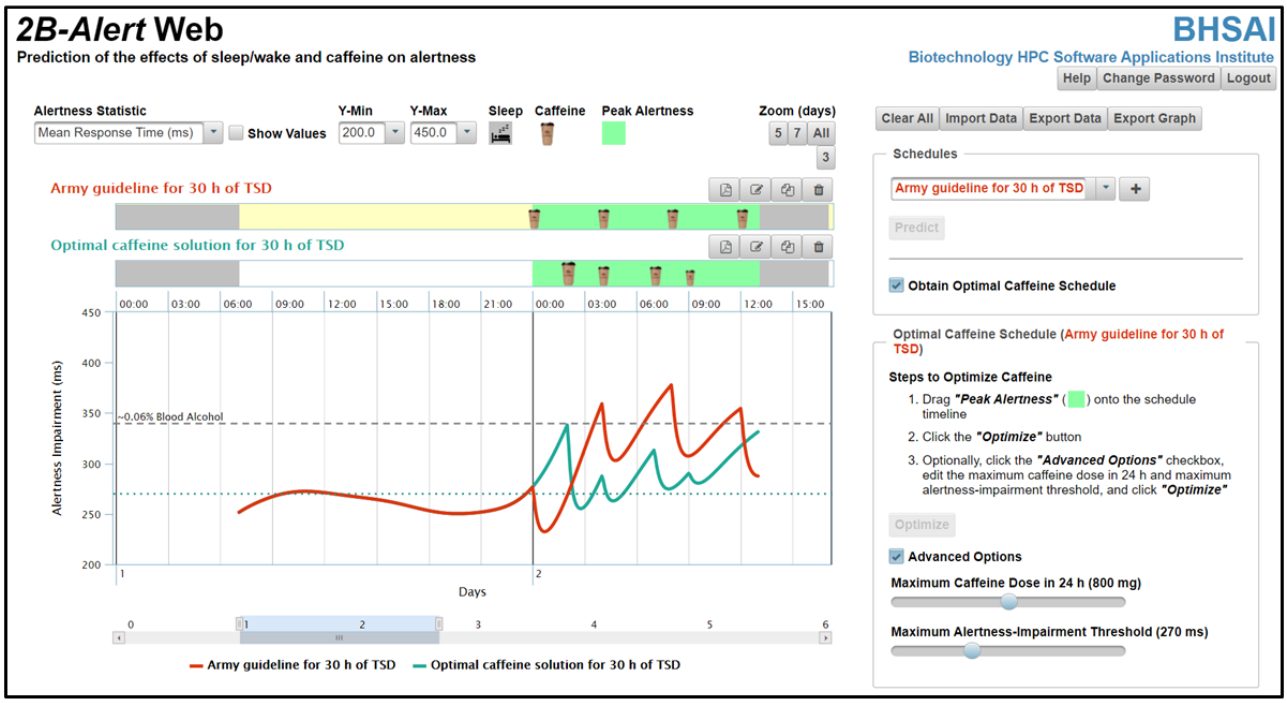
Condition	Army caffeine guideline	Sleep schedule	Peak-alertness period	Countermeasure recommendation	
				Army	*2B-Alert* Web
1. Sustained operations: 30 h[a] of total sleep deprivation	200 mg every 4 h as needed, starting at 00:00	N/A[b]	00:00-13:00 on day 2	800 mg in 4 doses (Figure 4A)	800 mg in 4 doses (Figure 4B)
2. Chronic sleep restriction: morning shift	200 mg upon awakening, 200 mg 4 h later	01:00-06:00 for 5 days	08:00-16:00 for 5 days	2000 mg in 10 doses (Figure S3A[c])	1900 mg in 13 doses (Figure S3B[c])
3. Chronic sleep restriction: evening shift	200 mg upon awakening, 200 mg 4 h later	01:00-06:00 for 5 days	15:00-23:00 for 5 days	2000 mg in 10 doses (Figure S5A[c])	1700 mg in 11 doses (Figure S5B[c])
4. Night shift with daytime sleep	200 mg at the beginning of the shift	20:00 to 22:00 and 10:00 to 15:00 for 5 days	00:00 to 08:00 for 5 days	1000 mg in 5 doses (Figure S7A[c])	1000 mg in 5 doses (Figure S7B[c])

[a]h: hours.

[b]N/A: not applicable.

[c]See Figures S3, S5, and S7 in Multimedia Appendix 1.

Figure 3. *2B-Alert* Web tool versus US Army caffeine recommendations for sustained operations (condition 1 in Table 1). Comparison of the effects of caffeine countermeasures as recommended by the US Army guidelines (top schedule) versus those automatically identified by the *2B-Alert* Web tool (bottom schedule) for 30 h of TSD with a user-defined peak alertness period ranging from midnight to 13:00 on day 2. Although neither solution was capable of maintaining alertness below the selected 270-ms impairment threshold with 800 mg of caffeine, the tool's solution (green line) avoided impairment levels surpassing the 0.06% blood alcohol concentration equivalence level. During the 13 h of peak alertness, it also reduced the mean alertness impairment by 56% and the peak alertness impairment by 36% (Table 2). h: hours; TSD: total sleep deprivation.



Alertness Improvement Metrics

To quantify the potential benefit of the caffeine-optimization algorithm, we used two metrics: mean alertness impairment and peak alertness impairment, each computed over the desired periods of peak alertness. The mean alertness impairment accounts for the overall impairment level across the peak alertness periods, whereas the peak alertness period captures acute levels of impairment over short periods of time during the peak alertness periods. For the mean RT PVT metric used in our simulations, we defined the mean alertness impairment (in ms) as the ratio of the area of the predicted alertness impairment above the selected maximum alertness-impairment threshold to the duration of the peak alertness periods. Similarly, we defined the peak alertness impairment (in ms) as the maximum mean RT over all periods of desired peak alertness minus the selected maximum alertness-impairment threshold. In addition, we computed the total time (in h) that the predicted alertness impairment was above the 0.06% BAC threshold during the peak alertness periods as a measure of time for which

an individual is at an increased risk of making a mistake because of deficits in alertness.

System Architecture

The overall software architecture of the updated *2B-Alert* Web tool was unchanged from the original release except for the overhaul of the graphical user interface and the use of updated versions of the underlying software technologies. We hosted the tool on an Apache Tomcat web server and provided access through a secure service over HTTP Secure. We used a three-tier architecture consisting of (1) a PostgreSQL database server, which stores user account information; (2) a controller, which provides access to the alertness-prediction model and caffeine-optimization algorithm and implements the functionalities required to create and manage multiple predictions and optimizations; and (3) an interaction-enabled user interface, which provides the ability to create schedules by dragging and dropping sleep, caffeine, and peak-alertness episodes onto the schedule timeline; show and hide plots; and dynamically zoom into and out of plots. The system runs without any plugins and is accessible through multiple web browsers, including Internet Explorer version 11, Chrome version 74, and Firefox version 67, or earlier versions thereof.

Access and Privacy

The *2B-Alert* Web tool is freely available to registered users through a secure web browser. Registration consists of users providing their name, email address, and affiliation, after which they receive a confirmation email with a username and password, which can be changed. The system erases all simulated schedules and results when the user logs out, providing user privacy and maintaining data confidentiality.

Results

Potential Benefits of the 2B-Alert Web Caffeine-Optimization Algorithm

To demonstrate the potential benefit of the *2B-Alert* Web caffeine-optimization algorithm, we compared the effects of caffeine countermeasures as recommended by the US Army guidelines [10] with those of the tool for the 4 conditions illustrated in Table 1. For the sustained-operations scenario in condition 1 (Figure 3), the guideline recommended an initial 200-mg caffeine dose at midnight, followed by 200 mg every 4 h thereafter (at 04:00, 08:00, and 12:00) for a total of 800 mg over a 12-h span (Figure 4A). Figure 3 shows the resulting alertness-impairment prediction plot in terms of the mean RT (red line). Figure 3 also shows the optimal caffeine schedule recommended by the tool and the associated alertness-impairment prediction plot (green line) for the same total amount of 800 mg of caffeine. Figure 4B, obtained by exporting the *Optimal caffeine solution for 30 h of TSD* schedule as a PDF, shows the numerical values of the optimal caffeine times and doses recommended by the tool. By suggesting caffeine interventions at different time intervals in varying doses of 100 mg, 200 mg, or 300 mg over a more compressed 7-h span during which time the circadian-mediated alertness deficits were greatest, the *2B-Alert* Web solution consistently kept impairment below the 0.06% BAC equivalence level during the entire peak-alertness period. In contrast, the US Army guideline solution was less effective, resulting in 3 time periods when alertness deficits exceeded the 0.06% BAC level for a total of 3.4 h above this threshold and impairment levels as high as 378 ms (270+108 ms) around 08:00 on *day 2* (Table 2 and Figure 3). To provide a context of the significance of this difference, an individual following the US Army guidelines would have a 2-fold increased risk of causing a traffic accident during the 3.4-h period compared with substantially lower increase in risk with the tool's recommendations. Nevertheless, for a maximum of 800 mg of caffeine, neither recommendation was capable of maintaining alertness levels below the maximum user-defined threshold of 270 ms for most of the 13-h period of peak alertness.

Figure 4. Sleep schedule, peak-alertness schedule, and caffeine recommendations for the results depicted in Figure 3: (A) US Army guidelines and (B) optimal caffeine solution automatically generated by the *2B-Alert* Web tool. Users export this information as PDF files by clicking on the left-most of the 4 buttons on the upper right-hand side of each schedule's timeline in Figure 3. h: hours; TSD: total sleep deprivation.

(A) ***Schedule:*** **Army guideline for 30 h of TSD**

Sleep				Peak Alertness				Caffeine		
Start		End		Start		End				
Day	Time	Day	Time	Day	Time	Day	Time	Day	Time	Dose (mg)
0	23:00	1	07:00	2	00:00	2	13:00	2	00:00	200
2	13:00	2	17:00					2	04:00	200
								2	08:00	200
								2	12:00	200

(B) ***Schedule:*** **Optimal caffeine solution for 30 h of TSD**

Sleep				Peak Alertness				Caffeine		
Start		End		Start		End				
Day	Time	Day	Time	Day	Time	Day	Time	Day	Time	Dose (mg)
0	23:00	1	07:00	2	00:00	2	13:00	2	02:00	300
2	13:00	2	17:00					2	04:00	200
								2	07:00	200
								2	09:00	100

Table 2. Predicted alertness impairment during the selected peak-alertness periods for the 4 simulated sleep-challenge conditions in using caffeine recommendations from the US Army guidelines and the *2B-Alert* Web optimization algorithm. Both the mean alertness impairment and the peak alertness impairment are computed considering impairment above the user-specified maximum alertness-impairment threshold. Although the mean impairment was averaged over the corresponding peak-alertness periods in the simulated days, the peak impairment values correspond to the maximum impairment over the same periods. The tool improved the mean alertness on average by 59% (SD 25%) and peak alertness by 45% (SD 18%). The time durations for which the predictions reached impairment levels above the 0.06% blood alcohol concentration (BAC) equivalent are provided to help with interpretation of the results.

Condition	Mean alertness impairment (ms)		Improvement (%)	Peak alertness impairment (ms)		Improvement (%)	Time above 0.06% BAC equivalent (h)	
	Army	*2B-Alert* Web		Army	*2B-Alert* Web		Army	*2B-Alert* Web
1	48	21	56	108	69	36	3.4	0.0
2	8	4	50	44	25	43	0.0	0.0
3	18	1	94	38	11	71	0.0	0.0
4	45	29	36	126	87	31	14.6	3.1

Caffeine Recommendations

Similarly, the *2B-Alert* Web caffeine recommendations for the sleep restriction conditions and the night shift with daytime sleep condition in Table 1 yielded superior results when compared with the US Army guidelines solutions. The tool's recommendations yielded larger improvements in mean alertness impairment and peak alertness impairment, while using the same or as much as 300 mg less caffeine than the US Army's recommendations (Tables 1 and 2). In condition 2, although the percentage improvement in mean alertness impairment was 50%, the absolute improvement averaged over the peak-alertness periods in the 5 simulated days was minimal (4 ms vs 8 ms). Given the limited amount of sleep deficit, the relatively narrow period of desired peak alertness (8 h per day), and the relatively large amount of consumed caffeine per day (400 mg), there were not enough degrees of freedom in this problem to demonstrate a considerable improvement of the optimization algorithm over the guidelines solution because it already provided near-optimal benefits. For related reasons, at times, the benefits achieved over peak-alertness periods came at the detriment of other wakefulness periods (eg, condition 3). Figures S2-S7 in Multimedia Appendix 1 provide detailed results for the simulations of conditions 2-4. Overall, our simulation results show that customized caffeine recommendations intended to enhance alertness during a specific time of the day consistently resulted in higher levels of alertness compared with suitable but generic recommendations.

Discussion

Principal Findings

We have enhanced the publicly available *2B-Alert* Web tool to automatically identify caffeine interventions that safely optimize alertness within selected time periods for any user-specified sleep and wake schedule. Users can use the tool to compare and contrast the effects of different sleep and wake and caffeine schedules on the alertness of a group of individuals and to automatically identify safe caffeine interventions to achieve and sustain desired alertness levels for desired times of day. To achieve this added capability, we integrated a well-validated mathematical model, which predicts alertness impairment levels based on the PVT as a function of time of day, sleep and wake schedule, and caffeine consumption, with a recently developed caffeine-optimization algorithm, which identifies the minimum amount of caffeine required to restore alertness levels, while imposing multiple operational and safety constraints (some of which can be modified by the user). The *2B-Alert* Web tool synthesizes decades of sleep-physiology and mathematical modeling research by the US Army into an open-access practical tool that should find value in research and operational communities.

To demonstrate the potential benefits of automatically identifying caffeine recommendations that safely optimize alertness for desired time periods, we simulated four conditions of military relevance reflective of sustained operations, restricted sleep (with morning and evening shifts), and night shift with daytime sleep (Table 1). Next, we compared the resulting alertness impairment levels by considering the US Army guidelines for the use of caffeine as a countermeasure to sleep deprivation [10] against those automatically obtained by using the newly implemented caffeine optimization capability [9] provided by the tool. The results show that the tool consistently provides solutions with greater improvements in mean alertness impairment (by 36% to 94%; mean 59%, SD 25%) and peak alertness impairment (by 31% to 71%; mean 45%, SD 18%), while using the same or as much as 300 mg less caffeine than the US Army's recommendations (Table 2). Although the extent of such benefits varies as a function of sleep and wake schedule, time and duration of desired peak alertness periods, acceptable alertness-impairment threshold, and maximum total amount of caffeine consumption, the results also demonstrate that the customized caffeine recommendations provided by this new capability are superior to generic, *one-size-fits-all* guidelines.

Limitations

The tool has limitations. First, the predictive models that form the core of the *2B-Alert* Web application have only been validated for healthy, young adults of military age. The extent to which the tool is applicable to an older, heterogeneous cohort is not known. Second, the tool's predictions of alertness impairment are in fact surrogates for PVT statistics. Hence, whether such predictions also serve as surrogates for other neurocognitive measures of performance impairment remains to be investigated. Third, all sleep and caffeine times are assumed to be for the same time zone. The ability to predict

alertness impairment involving transmeridian travel would require circadian-phase data across time zones, which we do not possess. Fourth, the *2B-Alert* Web tool does not account for the potential disruptive effect of caffeine when consumed soon before sleep episodes. In a future implementation, we could address this limitation by forbidding caffeine recommendations during certain times of the day. Fifth, the predictive models do not account for tolerance to caffeine. However, a sensitivity analysis of the caffeine-model component suggests that changes in model parameters to account for this effect have a very small impact on the predicted alertness [8]. Importantly, experimental observations indicate that the main determinant of the effect of caffeine on alertness is the impairment level, regardless of the caffeine tolerance level [15,16], which is consistent with the model predictions.

Sixth, the tool has not been prospectively tested in military or civilian settings such as the Marine Corps Officer Basic School or shiftwork at hospitals, and we estimated the benefits of the optimal caffeine schedules for the 4 simulation scenarios in Tables 1 and 2 using the tool's prediction models instead of using actual PVT measurements from an experimental study. However, the mathematical models at the core of the tool have been thoroughly validated in dozens of different sleep and caffeine-consumption conditions. For example, in one of the validation analyses, Ramakrishnan et al [7] showed that approximately 87% of the model predictions are within 2 SEs of the measured mean data, meaning that, 87% of the time, the model group-average predictions are statistically indistinguishable from the experimental data. Thus, we have reason to believe that the model will provide good predictions in conditions for which experimental data are not available, producing adequate guidance for sleep and caffeine schedules when it is not possible or desirable to conduct a study. Finally, we limited predictions to a maximum of 30 consecutive days to be able to provide solutions in a few (less than approximately 3) minutes. Nevertheless, the user can run longer scheduling scenarios by carefully breaking the problem into smaller chunks.

Conclusions

The new and unique capability to automatically identify caffeine interventions that safely optimize alertness transforms the *2B-Alert* Web tool from an alertness and performance *prediction* tool into an alertness and performance *enhancement* tool. This functionality not only results in peak alertness levels during desired wake periods but, by recommending the right caffeine doses at correct times, also maximizes the utility of caffeine. The *2B-Alert* Web tool provides caffeine schedules that use the least amount of caffeine to achieve a desired alertness level, thus reducing potential side effects of excessive caffeine use, or that yield *maximum* alertness levels for a given amount of caffeine, thus optimizing its recuperative effects.

Acknowledgments

The authors express their gratitude to Dr Tatsuya Oyama and Ms Maria Kuhrmann for editing the manuscript, Dr Bora Sul for providing constructive feedback, and the hundreds of study volunteers whose data we used to construct and validate our mathematical models over the years. This work is supported by the US Army Medical Research and Development Command under contract number W81XWH20C0031. This was not an industry-supported study. This work was sponsored by the Military Operational Medicine Program Area Directorate of the US Army Medical Research and Development Command, Fort Detrick, MD. The opinions and assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the US Army, the US Department of Defense, or the Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc. This paper has been approved for public release with unlimited distribution.

Authors' Contributions

JR, SR, and FGVL conceived the research. KK, LH, AF, and FGVL developed the software and performed the simulations. TJD and TJB provided user-perspective input to software workflow and graphical user interface. JR and FGVL wrote the manuscript, which was reviewed and edited by all authors.

Conflicts of Interest

JR, TJB and FGVL receive royalties for the licensing of the *2B-Alert* technology from Integrated Safety Support.

Multimedia Appendix 1
Supporting information.
[DOC File , 863 KB - jmir_v24i1e29595_app1.doc]

References

1. Biotechnology High Performance Computing Software Applications Institute. URL: https://2b-alert-web.bhsai.org [accessed 2021-01-13]
2. Reifman J, Kumar K, Wesensten NJ, Tountas NA, Balkin TJ, Ramakrishnan S. 2B-Alert Web: an open-access tool for predicting the effects of sleep/wake schedules and caffeine consumption on neurobehavioral performance. Sleep 2016 Dec 01;39(12):2157-2159 [FREE Full text] [doi: 10.5665/sleep.6318] [Medline: 27634801]

3. Rajaraman S, Gribok AV, Wesensten NJ, Balkin TJ, Reifman J. Individualized performance prediction of sleep-deprived individuals with the two-process model. J Appl Physiol (1985) 2008 Feb;104(2):459-468 [FREE Full text] [doi: 10.1152/japplphysiol.00877.2007] [Medline: 18079260]
4. Rajaraman S, Gribok AV, Wesensten NJ, Balkin TJ, Reifman J. An improved methodology for individualized performance prediction of sleep-deprived individuals with the two-process model. Sleep 2009 Oct;32(10):1377-1392 [FREE Full text] [doi: 10.1093/sleep/32.10.1377] [Medline: 19848366]
5. Rajdev P, Thorsley D, Rajaraman S, Rupp TL, Wesensten NJ, Balkin TJ, et al. A unified mathematical model to quantify performance impairment for both chronic sleep restriction and total sleep deprivation. J Theor Biol 2013 Aug 21;331:66-77 [FREE Full text] [doi: 10.1016/j.jtbi.2013.04.013] [Medline: 23623949]
6. Ramakrishnan S, Laxminarayan S, Wesensten NJ, Kamimori GH, Balkin TJ, Reifman J. Dose-dependent model of caffeine effects on human vigilance during total sleep deprivation. J Theor Biol 2014 Oct 07;358:11-24. [doi: 10.1016/j.jtbi.2014.05.017] [Medline: 24859426]
7. Ramakrishnan S, Wesensten NJ, Balkin TJ, Reifman J. A unified model of performance: validation of its predictions across different sleep/wake schedules. Sleep 2016 Jan 01;39(1):249-262 [FREE Full text] [doi: 10.5665/sleep.5358] [Medline: 26518594]
8. Ramakrishnan S, Wesensten NJ, Kamimori GH, Moon JE, Balkin TJ, Reifman J. A unified model of performance for predicting the effects of sleep and caffeine. Sleep 2016 Oct 01;39(10):1827-1841 [FREE Full text] [doi: 10.5665/sleep.6164] [Medline: 27397562]
9. Vital-Lopez FG, Ramakrishnan S, Doty TJ, Balkin TJ, Reifman J. Caffeine dosing strategies to optimize alertness during sleep loss. J Sleep Res 2018 Oct;27(5):e12711 [FREE Full text] [doi: 10.1111/jsr.12711] [Medline: 29808510]
10. Army Techniques Publication ATP 6-22.5 A Leader's Guide To Soldier Health And Fitness. United States: Army Techniques Publication; 2016.
11. Dawson D, Reid K. Fatigue, alcohol and performance impairment. Nature 1997 Jul 17;388(6639):235. [doi: 10.1038/40775] [Medline: 9230429]
12. Williamson AM, Feyer AM, Mattick RP, Friswell R, Finlay-Brown S. Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance. Accid Anal Prev 2001 May;33(3):313-326. [doi: 10.1016/s0001-4575(00)00045-2] [Medline: 11235793]
13. Borkenstein R. The Role of the Drinking Driver in Traffic Accidents (the Grand Rapids Study). Bloomington: Indiana University; 1974.
14. Martin TL, Solbeck PA, Mayers DJ, Langille RM, Buczek Y, Pelletier MR. A review of alcohol-impaired driving: the role of blood alcohol concentration and complexity of the driving task. J Forensic Sci 2013 Sep;58(5):1238-1250. [doi: 10.1111/1556-4029.12227] [Medline: 23879433]
15. Landolt H, Rétey JV, Adam M. Reduced neurobehavioral impairment from sleep deprivation in older adults: contribution of adenosinergic mechanisms. Front Neurol 2012;3:62 [FREE Full text] [doi: 10.3389/fneur.2012.00062] [Medline: 22557989]
16. Rétey JV, Adam M, Gottselig JM, Khatami R, Dürr R, Achermann P, et al. Adenosinergic mechanisms contribute to individual differences in sleep deprivation-induced changes in neurobehavioral function and brain rhythmic activity. J Neurosci 2006 Oct 11;26(41):10472-10479 [FREE Full text] [doi: 10.1523/JNEUROSCI.1538-06.2006] [Medline: 17035531]
17. Caffeine for the Sustainment of Mental Task Performance Formulations for Military Operations. Washington, DC: The National Academies Press; 2001.

Abbreviations

BAC: blood alcohol concentration
h: hours
PVT: psychomotor vigilance test
RT: response time
TSD: total sleep deprivation

Edited by R Kukafka; submitted 14.04.21; peer-reviewed by R Marshall, YH Lin, YS Huang; comments to author 24.09.21; revised version received 12.10.21; accepted 15.11.21; published 27.01.22.

Please cite as:
Reifman J, Kumar K, Hartman L, Frock A, Doty TJ, Balkin TJ, Ramakrishnan S, Vital-Lopez FG
2B-Alert Web 2.0, an Open-Access Tool for Predicting Alertness and Optimizing the Benefits of Caffeine: Utility Study
J Med Internet Res 2022;24(1):e29595
URL: https://www.jmir.org/2022/1/e29595
doi:10.2196/29595
PMID:35084336

Original Paper

Credit and Recognition for Contributions to Data-Sharing Platforms Among Cohort Holders and Platform Developers in Europe: Interview Study

Thijs Devriendt[1], MSc; Pascal Borry[1], PhD; Mahsa Shabani[2], PhD

[1]Department of Public Health and Primary Care, Faculty of Medicine, KU Leuven, Leuven, Belgium

[2]Metamedica, Faculty of Law and Criminology, UGent, Gent, Belgium

Corresponding Author:
Thijs Devriendt, MSc
Department of Public Health and Primary Care
Faculty of Medicine
KU Leuven
Kapucijnenvoer 35 blok d box 7001
Leuven, 3000
Belgium
Phone: 32 16193725
Email: thijs.devriendt@kuleuven.be

Abstract

Background: The European Commission is funding projects that aim to establish data-sharing platforms. These platforms are envisioned to enhance and facilitate the international sharing of cohort data. Nevertheless, broad data sharing may be restricted by the lack of adequate recognition for those who share data.

Objective: The aim of this study is to describe in depth the concerns about acquiring credit for data sharing within epidemiological research.

Methods: A total of 17 participants linked to European Union–funded data-sharing platforms were recruited for a semistructured interview. Transcripts were analyzed using inductive content analysis.

Results: Interviewees argued that data sharing within international projects could challenge authorship guidelines in multiple ways. Some respondents considered that the acquisition of credit for articles with extensive author lists could be problematic in some instances, such as for junior researchers. In addition, universities may be critical of researchers who share data more often than leading research. Some considered that the evaluation system undervalues data generators and specialists. Respondents generally looked favorably upon alternatives to the current evaluation system to potentially ameliorate these issues.

Conclusions: The evaluation system might impede data sharing because it mainly focuses on first and last authorship and undervalues the contributor's work. Further movement of crediting models toward contributorship could potentially address this issue. Appropriate crediting mechanisms that are better aligned with the way science ought to be conducted in the future need to be developed.

(J Med Internet Res 2022;24(1):e25983) doi:10.2196/25983

KEYWORDS

information dissemination; qualitative research; ethics; database management systems; cohort studies; science policy; incentives; rewards

Introduction

Background

Data sharing in science maximizes the utility and impact of cohort data and, therefore, contributes to improving clinical practice and public health. In this paper, data sharing will be considered as "making data available to other researchers for carrying out scientific analyses." This definition encompasses not only the transfer of data out of the institution but also the remote use of data from outside the institution (eg, through federated analyses). Data sharing also increases the accountability of researchers as others can rerun any analyses that are published or use different methodologies to answer the

same research question to replicate findings [1]. Furthermore, data sets may be used to explore secondary hypotheses, for meta-analyses within systematic reviews, or for educational purposes [1]. Despite these advantages, academics from diverse fields have emphasized that broad data sharing is impeded by the lack of incentives for those who share data [2-5]. In this context, incentives for researchers are often understood as interventions that can stimulate researchers to engage in particular behavior—here, open science practices. They may compensate for the lack of recognition or rewards for those who share their data. Empirical studies on data sharing indicate that recognition and credit can be major concerns of researchers [2,6-9]. The salience of granting due credit was also underlined when the WorldWide Antimalarial Resistance Network data platform became operational. Within the WorldWide Antimalarial Resistance Network, many of those who generated data within their local contexts (hereafter referred to as *data generators*) only agreed to contribute after being offered coauthorship on papers for downstream data use [10].

Offering coauthorship has become a commonly used means of crediting data generators. Through this mechanism, the rising prevalence of multi-cohort, multicenter studies has also been driving authorship inflation over time and has contributed to the phenomenon of hyperauthorship [11,12]. The term hyperauthorship was coined in 2001 by Cronin [11], who used it to refer to articles with "massive coauthorship levels" [12]. In this paper, hyperauthorship will be considered as *>100 authors*. However, the validity of traditional crediting models has come under mounting pressure as it raises questions about credit misconduct, accountability of authors, and the undue influence of hyperauthorship on popular metrics of scientific productivity [11-13]. Multiple scholars have suggested that hyperauthorship has the double effect of multiplying the credit attributed to the production of knowledge and fragmenting the responsibility among all authors [13,14]. At the same time, the academic reputation of all the authors may still be affected by the errors or mistakes of 1 author, albeit the degree of impact depends on the author's position [15]. A recent survey found that 46.6% and 37.9% of researchers reported having encountered questionable or unethical behavior with regards to author naming and ordering, respectively [16]. The highest degrees of dissatisfaction with author order, ghost authorship, and gift authorship are found among early-career academics and within health sciences [16,17].

In response to both challenges, namely academic recognition and accountability, alternative recognition approaches have been proposed, such as through designations that elucidate the contributions to scientific work in a more granular fashion [18,19]. At the end of the 1990s, Rennie et al [14] argued in favor of abandoning authorship and moving toward contributorship, where all contributions would be disclosed. Contributorship has the advantage of being able to more adequately recognize *specialist* authors (ie, those authors who contribute through 1 or 2 roles within research projects). In parallel, expert groups providing advice to policy makers are advocating moving away from commonly used and abused metrics, such as H-index and the journal impact factor, toward so-called *responsible* metrics [20]. The San Francisco Declaration on Research Assessment and the Leiden Manifesto for Research Metrics take similar positions on the development of novel indicators of scientific productivity and underscore the need to reconsider the relationship between quantitative (indicators) and qualitative (judgment) aspects of evaluation [21,22]. In recent years, these thoughts have started to gain traction and have become incorporated within policy documents on the open science agenda, such as documents of the working expert groups and mutual learning exercises of the European Commission [20,23-28]. Central to these documents is the understanding that incentives partly drive the behavior of scientists and that the installation of adequate incentives and rewards is essential to achieve open science practices (including greater data sharing).

Objective

Although perspective pieces on these issues are plentiful, few qualitative studies have explored the views of scientists on potential modifications of the recognition system. Colledge et al [29] recorded the experiences of biobank stakeholders with publication credit in collaborative research. Pinel [30,31] described the dynamics of competition, credit, and financial resources in collaborations in epigenetics based on an ethnographic study. Sauerman et al [32] went further and conducted a survey on researchers' responses on the perceived usefulness of contributorship statements for evaluation purposes, in which qualitative data on researcher perspectives were also collected. Empirical studies on data sharing and prior experiences with contributing to data-sharing platforms indicate that credit is an important factor. Therefore, our study explores the perspectives of cohort holders and platform developers on credit for data sharing and aims to describe concerns on this issue in detail within the context of cohort research. Furthermore, we asked researchers about their views on potential alterations to the reward system and gave them the opportunity to make suggestions for changes.

Methods

Overview

Qualitative methods were used to explore the views and opinions of cohort holders and platform developers. Cohort holders were understood to be those involved in generating the cohorts and platform developers to be those involved in designing data-sharing platforms. Most participants could be classified as cohort holders or both cohort holders and platform developers. They were recruited using a purposive sampling strategy that explored 3 European projects creating data-sharing platforms (euCanSHare [a European Union-Canada joint infrastructure for next-generation multistudy heart research], Common Infrastructure for National Cohorts in Europe, Canada, and Africa, and EUCAN-Connect). These 3 projects are funded under the same Horizon 2020 call. Contact persons for cohorts within the euCanSHare consortium were identified via a list of names acquired from the project manager and other lists found on the web. Within euCanSHare, nearly all European cohorts were contacted, except those for whom the principal investigator had passed away or where multiple cohorts were managed by the same center. Names of contact persons from the Common

Infrastructure for National Cohorts in Europe, Canada, and Africa and EUCAN-Connect consortium were acquired by contacting the project coordinators or by querying the participating cohorts in databases and registering the first or last authors on cohort profiles or recent articles. Potential interviewees were contacted via email. Interviews were semistructured in nature and were conducted by TD using a semistructured interview guide. The interview guide was updated after conducting the first interview. Interviews were audio recorded, transcribed verbatim, deidentified, and analyzed using inductive content analysis in which content categories were derived from the data rather than being predetermined [33-35]. Transcripts were coded into narrow content categories using NVivo 12 software (QSR International). Subsequently, categories were compared, revised, and broadened through iterative exploration of the excerpts, the coding scheme, and the original transcripts. Upon completion, the coding scheme and extracts were checked by MS for consistency and rationale. The coding scheme was further revised upon receiving feedback. All participants signed written informed consent. The interviews were conducted on the web and recorded. In total, 17 interviews were conducted: of these, 13 (76%) with cohorts affiliated with euCanSHare and 4 (24%) with the other data-sharing platforms. The study was approved by the social and societal ethics committee at Katholieke Universiteit Leuven (G-2018 10 1348).

Terminology

Cohort holders are those researchers who are apt at deciding on the use of cohort data in research projects.

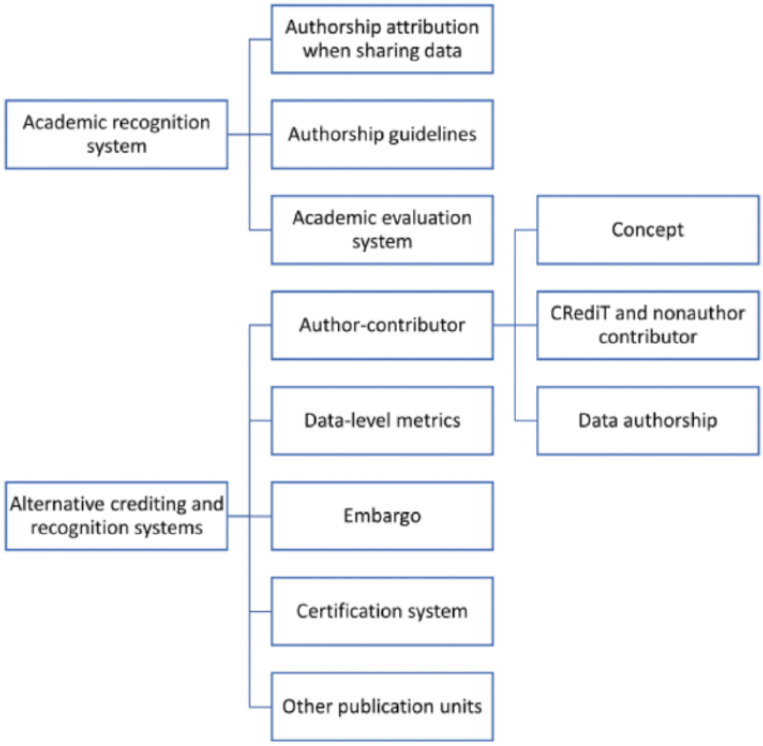
Data generators are those researchers who have been involved in *any* aspect of the production of cohort data, such as conception, data collection, or quality assurance.

Data contributors are those researchers who contribute data to projects to be used by other scientists within the framework of multicohort, multicenter studies.

Results

The subsections of the interview transcripts pertaining to credit were classified into the following main categories: (1) authorship attribution when sharing data, (2) authorship guidelines and data sharing, (3) authorship and the academic evaluation system, and (4) alternative crediting and recognition mechanisms. The tree diagram for these categories is displayed in Figure 1.

Figure 1. Coding scheme for the interview data. CRediT: contributor roles taxonomy.



Authorship Attribution When Sharing Data

Most interviewees communicated that the sharing of data was usually rewarded with authorship. A minority of interviewees considered that they or persons within their research team did not need to be authors when sharing data but required formal citation (eg, through cohort profiles) or acknowledgment of the cohort. Most respondents stated that they only shared data when collaborating on downstream analyses and that the involvement of researchers affiliated with the study was required:

> *In general, [the researchers collecting] the cohorts are interested in sharing their data and participating in the analysis [...] They have created their cohorts for research and [...] they want to be part of the research and therefore authorship is important [...] In general, [researchers will not share if they do not get the opportunity to participate in the analysis]. It varies very much from center to center.* [Interviewee 1]

Interviewees explained that, within research consortia, various modes of working could be adopted. Within some consortia, all participating research teams are able to suggest proposals for analyses. First, the scientific validity of a research proposal would be reviewed by a central committee associated with the consortium. If it passed this stage, it would be subsequently distributed to all groups in the consortium whose data are to be used in the requested analysis. All these groups would then be offered the opportunity to collaborate in the research project and would have to explicitly confirm their participation. Several rounds of drafting and circulation of the data analysis plan among participants would follow, which allow active discussion of methodology and technical details. In contrast, within larger consortia, especially genetic research consortia, respondents argued that it might not be possible to involve all data contributors in the drafting of the data analysis plan. This created some frustrations among interviewees, as they would not be truly involved in the research:

> *Manuscript leaders should try to increase transparency over all over the work that they are doing because if I receive an advanced draft on a very complicated genetic issue with forty tables only in the supplementary material and maybe fifteen days to provide the feedback, this means: Please do not say anything, just check the affiliation [...] You cannot have a strict law or rule but it is a gentlemen's agreement.* [Interviewee 7]

Respondents argued that often 1 or 2 persons per cohort would be allowed to become coauthors on resultant manuscripts. The need to limit the number of coauthors on these articles was mentioned by most respondents, and it was frequently referred to as a balance between the number of authors and the due credit to researchers for data production and sharing.

Authorship Guidelines and Data Sharing

The interviewees indicated that receiving approval and comments from all coauthors was not straightforward. A minority of interviewees brought up that, within consortia, the discussion had arisen about whether researchers who have never responded to emails requesting commentary should still be listed as coauthors:

> *You send them the manuscript with the complete analysis, written and ready to be submitted, asking them to read and make comments. There are people that in five, six years, have never answered. Not just to comment but to say: "Okay, I received the manuscript and this is okay for me."* [Interviewee 13]

Of the 17 respondents, 1 (6%) respondent argued that for older cohorts, the researchers involved in the original data collection might not be available anymore, for example, because of retirement. This means that persons managing that cohort might not have been involved in the acquisition of the original data. This makes it more complicated for researchers to conform with the first criterion of the International Committee for Medical Journal Editors (ICMJE) guidelines, which stipulates that authors need to have delivered "substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data for the work" [36]. Nevertheless, interviewees described experiences where cohort holders simply requested coauthorship regardless of compliance with the authorship guidelines.

Of the 17 respondents, 1 (6%) respondent pointed out that limiting the number of coauthors per cohort might mean that some researchers who have contributed are not recognized. At the same time, if a center contributes multiple cohorts, more names might be put on research articles than those that would qualify for authorship:

> *It is also maybe a coincidence whose name will be on this list. For example, if we have five cohorts in a collaboration, then we can put ten names on the article and the first two or three will be someone that will actually have something to do with the data but the last five or seven names will be like: pick one.* [Interviewee 11]

Respondents generally considered that for multi-authored papers, the authors who contributed data could only be accountable for their contributions rather than for the whole content of the paper or papers. This means that authors must be able to vouch for the integrity of the data and the proper conduct of any analyses that were performed locally. They can also review the manuscript for overinterpretations, errors in the background information, and an accurate description of the cohort itself. Respondents mentioned that, apart from their own contributions, one should simply put trust in others. Notably, a respondent referred to the need to have data contributors as coauthors, as the group leading the analysis cannot take responsibility for the proper handling of the data if they do not understand the particularities of the data.

Interviewees considered that data contributors might not be able to exert sufficient influence to correct potential errors in the manuscript or request additional analyses. Mistakes in the reporting of the cohort might reveal that data were incorrectly handled; however, redoing the analysis and reworking the tables could be very time consuming and substantially delay publication. Researchers might be disincentivized to comment if they anticipate that their views will not be duly considered or if they feel that others might blame them for causing time delays. Those leading the analysis might give little weight to comments if these comments are not shared by other researchers.

Authorship and the Academic Evaluation System

The aforementioned difficulties with claiming authorship and responsibility might create problems for acquiring recognition: if authorship is considered unjustified or too easily obtained, it loses its value. Most respondents argued that, for evaluation purposes, the first and last authors are considered to be most important, with little focus on other positions. Although some respondents stated that they were not considered for evaluation purposes at all, others expressed confidence in the first, last, and middle authorship being appropriately weighted. Of the 17 interviewees, 3 (18%) reported experiences with funders or evaluation panels being critical of a large number of papers with extensive author lists and that these funders and evaluations panels would not recognize the value in such authorship when no specific *evidential* contributions were reported:

> *When I tried to become a research director, it was good that I valorized the data and the fact that I had many publications thanks to this, contributed to me to go through the first step so my file was selected [and] I could defend my project in front of the jury. But once I was [there] the [fact] that I was mainly sharing data and not conducting my own scientific projects was a very bad thing. [...] I had difficulties defending [these actions] because it seemed that I was just a platform and not really directing some important research.* [Interviewee 16]

Interviewees raised concerns over the lack of recognition for data generators and researchers with specialized roles within our current system. According to these interviewees, the recognition system in academia is geared primarily toward clinical researchers and less toward other types of researchers such as statisticians and data generators:

> *Researchers who contribute through data sharing or the creation of data that then others use, I think that needs to be acknowledged and is not really acknowledged in the middle of a paper. [...] There is a need to be able to recognize people who are doing more of the hands-on work, the less research-type work, they are building the resource. That is not recognized to the same extent as then using the data and scientific output. [...] [I would prefer] if funders were more willing to fund support persons rather than just research, post-docs and clinical [persons] who all have the incentive to publish in the first and last authorship role.* [Interviewee 12]

Respondents argued that junior researchers are disadvantaged by this system, as they are in need of publications in which they are the first author. They are often not experienced enough to lead the analysis of a large number of cohorts, and when collaborating, they would only be in positions where their work remains largely unrecognized. In addition, junior researchers might intend to perform a similar analysis on local cohort data as planned by the consortium, which would threaten their publication opportunities. At the same time, it is often the junior researchers who might have dedicated a substantial part of their time to collecting data or performing quality assurance.

Despite the difficulties surrounding authorship in terms of attribution, accountability, and credit, most interviewees generally stated that they were satisfied with the ICMJE or Vancouver guidelines and considered that they form an appropriate basis for discussing authorship disputes. Many respondents held pragmatic views on offering authorship, which, they argued, recognizes scientific work or stimulates data sharing to the benefit of science. Among the interviewees who suggested changes, two opposing views could be identified: (1) authorship should be more inclusive of data generators and those in supporting roles and (2) authorship should be only for those who have actively and intellectually contributed to the analysis. Adherents to the former view based their argumentation on the inequity of unrecognized work by specialists or data generators. Adherents to the latter view based their argumentation primarily on authorship losing its value or going against what it used to be.

Alternative Crediting and Recognition Mechanisms

Contributor Roles Taxonomy and Data Authorship

Researchers were asked to state their opinions on multiple proposals to reform the academic recognition system. Most respondents declared their willingness to explore alternatives to authorship, such as the specification of contributorship roles. Respondents considered that this could resolve the concerns around accountability to display that one has contributed substantially to articles with extensive author lists and allow better recognition of specialized researchers, such as those contributing to only 1 or 2 tasks within large collaborations. The principal downsides perceived by participants were related to increased bureaucracy and complexity for evaluation purposes. Some argued that contributor roles taxonomy (CRediT) could disincentivize credit misconduct by forcing authors to declare their contributions, whereas others argued that it does not make a difference, as boxes may be very easily ticked without subsequent verification. Of the 17 respondents, 1 (6%) respondent stated that when data were generated, the name of the data generator was added to the code. This concept of *data authorship* can then make persons eligible for authorship when the said data are reused. Although respondents were generally open to alternatives, many explained that their views would depend on the manner in which these systems are eventually used for evaluation purposes:

> *[CRediT] is an interesting concept. It has some potential. On the other hand, I would not like to see things getting to complicated, too long lists, different alternatives. There are then differences in how people in different countries interpret those boxes [with different roles]. It might be good but it should be kept relatively short and relatively clear.* [Interviewee 2]

Nevertheless, some respondents considered that the lack of recognition of these categories by universities and funders could create barriers to contributing to research. Researchers might be dissuaded from engaging in activities that do not provide credit:

> *The introduction of non-author contributors was a very big step forward. Now the problem is within the scientific community, some universities whose actions are telling [you] that your work is recognized only if you are an author. They create barriers for the researchers to contribute to research.* [Interviewee 1]

Certification Systems, Data-Level Metrics, and Alternative Publication Units

In addition to these designations related to publications, interviewees suggested several systems to confer greater credit upon those who generate valuable data. Of the 17 respondents, 1 (6%) respondent argued that recognition systems could be set up by funders or institutions to recognize the broader merit of cohort studies, such as the way in which it actively involves participants, the quality of the data collection, and the way in which it disseminates relevant expertise. Such a certification system could include evaluating study practices according to internationally recognized criteria.

Approximately 18% (3/17) of participants argued in favor of the implementation of data-level metrics related primarily to either data quality or data use. Of the 17 interviewees, 1 (6%) interviewee argued that funders should shift toward indicators of high-quality data earlier in the process of data generation, which compares data with a fixed set of quality indicators. Other interviewees considered such metrics to be highly context dependent, not always useful for evaluation purposes, or to potentially instigate undesirable modes of competition (eg, by stimulating untruthful behavior). Some respondents considered that data use metrics could be relevant for decision-making. For example, they could be used to create *data set profiles* displaying collaborative sharing, data contribution, and sharing for educational and training purposes or partnerships with institutes in other countries.

Of the 17 participants, 1 (6%) participant argued in favor of publishing other publication units, such as the data dictionary and study protocol, in dedicated journals. Data generators are then to be evaluated based on the outputs that are more directly related to their work.

Discussion

Principal Findings

Our study indicates that granting authorship is commonly used to reward those who produce and share data. However, there is substantial divergence in the degree of involvement, from being actively involved in drafting data analysis plans or performing local analyses to sharing data without subsequent involvement in downstream analysis. When sharing data, the number of authors per cohort may be restricted to keep author lists more concise. Multiple challenges can arise with regards to the application of the ICMJE guidelines, such as when collaborators do not respond to requests for revision of manuscripts or when researchers who collected the data have already retired. It can be difficult to acquire credit for papers with extensive author lists if evaluation procedures for tenure or funding acquisition put excessive focus on first or last authorship. Those scientists in roles that are more prone to end up in middle author positions may be disadvantaged. Furthermore, researchers may be, in some instances, criticized for spending too much time on sharing data rather than leading the research themselves. Researchers look upon alternative crediting systems as potential solutions for addressing concerns over accountability or credit but refrain from supporting these endeavors wholeheartedly when the details of their implementation and its consequences remain unclear.

Comparison With Previous Work

To some extent, academic credit influences researchers' decisions to share data and the manner in which the sharing takes place. As interviewees indicated, there is an opportunity cost associated with spending time on data sharing, as this time could be spent on leading research projects instead. If data sharing is supported by a dedicated staff (eg, for data harmonization) or if administrative procedures are centralized, this factor would become less important. The choice between the 2 activities will then be partly influenced by the perception of rewards for data sharing in comparison with data analysis. If evaluation systems, formally through metrics or informally through qualitative evaluation, are tailored to primarily evaluate data analysis, the use of (local) resources could be considered more rewarding than contributing to collaborative work. Tornetta et al [37] described this dilemma between publishing smaller, less clinically relevant studies based on local data or contributing to large research groups and potentially jeopardizing one's career. In contrast, Pinel [30,31] emphasizes the collaborative imperative in contemporary medical science but argues that engaging in collaborative work (eg, by joining consortia) is accompanied by the monopolization of that data and forced exclusion of outsiders. In addition to gaining credit for data sharing, this suggests that data sharing may still serve as the buy-in mechanism of acquiring access to high-quality data of others (either through reciprocal data sharing one on one or in consortia).

Multiple scholars have voiced concerns over recognition systems that may disincentivize scientists from contributing to multicenter, multidisciplinary research [13,37-39]. A recent analysis of guidelines for promotion within biomedical faculties revealed that few universities mention middle authorship or data sharing as valuable criteria [40]. Within biomedical sciences, special significance is often attached to first and last authorship, with the exception of large collaborations where more focus is on the ones in the beginning (eg, first and second) and the end of the authorship list (last and second-to-last) [41]. However, Mongeon et al [42] found that, for biomedical sciences and clinical medicine, the fraction of middle authors in an article is not connected to the total number of coauthors, implying that authorship inflation is not only because of an increase in the number of specialized contributions (eg, those who only contribute data). At the same time, clinical research and biomedical sciences possess a great degree of division of labor, with many researchers contributing to only 1 or 2 specific tasks [13,43]. These results suggest that focusing exclusively on first or last authorship for evaluation purposes could be detrimental to those authors who contributed to many tasks but do not take the first or last position and those who make many specialized contributions to various articles. Mazumdar et al [44] argued that the focus on first authorship is suitable for evaluating the work of small research groups, yet that this can be a disadvantage to typical team scientists, such as biostatisticians. Researchers may consider that certain types of contributions tend to be undervalued, such as technical analysis, model creation, and statistical analysis [45]. Through ethnographic studies, Pinel [30,31] describes the process in (epigenetic) cohort research, where PhD candidates and research nurses (and to some extent, postdoc holders) are involved in the *invisible* processes of data production, data curation, and *housekeeping jobs* of the laboratory. Pinel et al [46] criticize that engaging in these tasks has low prestige and is perceived as taking away from the *real science* and writing of publications. As suggested during the interviews, last authorship is considered particularly important for tenure as it emphasizes the independence of the researcher. Soares [47] has criticized the practice of attributing too much weight to independence as a criterion and misrecognizing the existing interdependence between researchers within collaborative work. The fact that many proposals for alternative crediting systems capitalize upon the

perceived lack of recognition for specialized contributors merely underscores the salience of these issues [48].

One such crediting system that has substantially influenced authorship and gained popularity is contributorship. In the late 1990s, Rennie et al [14] proposed abandoning authorship and adopting contributorship in response to (1) changes in authorship practices and the way scientific research is conducted, (2) concerns about the dilution of accountability of authors, and (3) as an attempt to recognize all the authors who contributed more equitably. Contributors would simply be those who have contributed substantially and agree to take responsibility for the contribution without having to be involved in writing or revising the manuscript. In addition, guarantorship would be introduced, which refers to those who have contributed substantially to the article and have made efforts to organize, oversee, and double-check the integrity of the work [14]. Although its adoption was initially limited, it has become increasingly common in journals in the field of biomedicine over time. In 2015, the CRediT was introduced to standardize contributorship statements [49]. The use of CRediT has become more common in recent years and is anticipated to increase substantially in the near future [50]. The use of CRediT is argued to have advantages such as enabling meta-research programs, reducing credit misconduct, facilitating cross-disciplinary collaboration, and appropriately recognizing specialist roles and the contribution of software or data [18]. Within this interview study, researchers looked positively upon contributorship to elucidate who has contributed what to the research article, facilitate acquiring credit for specialists, and designate accountability. Along the same lines, Sauermann and Haeussler [32] found in their survey that >90% of researchers consider that contributorship statements add at least some additional information, with 40% seeing considerable or much additional information. They also found that 45% of respondents favor authorship order for the evaluation of researchers, whereas 18% would give greater weight to contributorship statements, with junior scientists attributing comparatively more weight to contributorship statements than senior scientists [32]. The lack of necessity, increased bureaucracy, divergent interpretation of roles across groups, lack of universal adoption, difficulty in accessing aggregate data, and the inability to divide teamwork were touched upon at least once during the interviews. The supplementary data of the survey by Sauermann and Haeussler [32] revealed similar concerns of scientists, with the addition of lack of detail (eg, weight or categories), risk of introducing conflict, and lack of experience in using alternative schemes for evaluation. Despite these potential drawbacks, contributorship could still evolve over time. For example, some scholars argued in favor of contributorship statements indicating both absolute (ie, form) and relative contributions (ie, weightage) [32,41,51]. CRediT could also evolve as science progresses or could be tailored to become discipline specific [50]. McNutt et al [52] recommended that authorship roles be embedded within author metadata to allow indexing and abstracting, which would facilitate their use for evaluation purposes.

The opposing views put forward by interviewees that authorship should (1) reaffirm its more *traditional* meaning by encompassing only those who have contributed intellectually to the analysis or (2) be expanded to include data generators and those in supporting roles should be understood from the history and evolution of data sharing within the ICMJE guidelines. Before the 2000s, the first criterion of the ICMJE guidelines stated that acquiring authorship required "substantial contributions to (1) conception and design, or analysis and interpretation of data" whereas explicitly stating that the collection of data itself did not justify authorship [53]. In May 2000, the guidelines were revised, and the acquisition of data was added to the first criterion [54]. In the version of August 2013, the statement that the collection of data itself does not merit authorship was omitted [55]. Finally, in 2017, the guidelines stipulated that data generation and sharing deserve substantial credit; data generators must be given the opportunity to collaborate; and if this is impossible, unpractical, or undesirable, the efforts of those who shared data should still be recognized [56]. A comparable reversal can be observed in the notion of accountability. Before the 2000s, all authors were responsible for the entirety of the work [53]. In May 2000, authors became responsible for "appropriate portions of the text" [54]. As of August 2013, authors are accountable for the parts of the work that they had performed and for which they needed confidence in the contributions of others [55]. Thus, the conception of authorship-worthy contributions has broadened to encompass data sharing, whereas the notion of accountability has narrowed to resolve the problems created by far-reaching specializations in science. In this sense, authorship has edged closer to contributorship over the past 20 years. With this understanding, the normative positions taken by interviewees on authorship simply reflect support for or opposition to this shift. It remains uncertain what the precise connection will be between authorship and contributorship in the future. As Smith and Master [43] have argued, perhaps contributorship will not replace authorship but will become the basis for it.

Finally, any reformation of the academic evaluation system should consider how research pipelines will commonly be organized in the future. A true shift from local to big data epidemiology could entail fundamental policy changes, such as more emphasis on centrally driven initiatives, fewer population studies being supported, or dedicated funding streams for data analysis [57,58]. Therefore, commentators have raised questions about how the academic interests of original investigators and data analysts will be balanced in the future and how participation in multicenter, international epidemiological studies would be seen in terms of career advancement [57,58]. Some have expressed concerns that the shift toward big data epidemiology will preclude junior researchers from opportunities in first-author articles or that credit and resources will be accumulated (ie, the Matthew effect) in fewer hands [59,60]. Contributions to data-sharing platforms are not disconnected from this social reality of academia, where academic competition and the acquisition of credit partly influence researchers' behavior. Therefore, it remains essential that robust science policies, including sound recognition systems, can support these platforms to ensure that their full potential is realized.

Limitations

Many of the recruited interviewees actively participated in collaborative research within the existing consortia. Therefore, positive experiences with such modes of working may have steered the opinions in favor of active collaboration. With cohorts outside of euCanSHare, interviewees held more heterogeneous views, especially when data generation was funded for the principal purpose of sharing data broadly, when already having experiences with progressive evaluation criteria, or when faced with higher degrees of centralization of research governance. As such, the discussion on incentives for data sharing is likely to be *highly* context dependent, with cultural differences existing even within different branches of cohort research and researchers being subject to different national, institutional, or departmental regulations and rules. This interview study probed the views of researchers on novel crediting systems while providing a minimal explanation of recent evolutions in terms of authorship, contributorship, other crediting systems, or their place within the open science agenda. Therefore, engaging with researchers in deliberation, where the flaws of science policy, the history of particular initiatives, the strengths and shortcomings, and policy evolutions in open science are explained, may still shift opinions.

Conclusions

The acquisition of credit influences the degree and mode of sharing. The evaluation system might impede data sharing through the undervaluation of scientific work as contributors. The challenges posed by an increase in collaborative science led to the suggestion of adopting contributorship. Subsequently, the authorship model has changed positions on accountability and credit for data production and sharing, whereas contributorship has become standardized and increasingly common. These evolutions underscore the salience of providing due credit to data generators and team scientists. Contributorship statements may be able to better support recognition systems for data-intensive and collaborative science. Contributorship may further evolve by becoming discipline specific, incorporating relative weight or allowing indexation. Researchers might accept further movement toward contributorship, although the final test will be its utility for evaluation purposes.

Acknowledgments

This paper is part of a project that has received funding from the European Union's Horizon 2020 research and innovation program under grant 825903. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors wish to sincerely thank all the interviewees for dedicating their time to participate and share their views, opinions, and ideas and enable the conducting of this study.

Authors' Contributions

TD and MS conceptualized the study. TD performed the interviews, background literature search, and data analysis. MS verified the coding scheme and suggested alterations. TD wrote the first draft of the manuscript. MS and PB reviewed the draft and provided feedback. TD further refined the draft. All the authors reviewed the manuscript.

Conflicts of Interest

None declared.

References

1. Barbui C, Gureje O, Puschner B, Patten S, Thornicroft G. Implementing a data sharing culture. Epidemiol Psychiatr Sci 2016 Aug;25(4):289-290 [FREE Full text] [doi: 10.1017/S2045796016000330] [Medline: 27150624]
2. Dallmeier-Tiessen S, Darby R, Gitmans K, Lambert S, Matthews B, Mele S, et al. Enabling sharing and reuse of scientific data. New Rev Inf Netw 2014 Apr 18;19(1):16-43. [doi: 10.1080/13614576.2014.883936]
3. Ali-Khan SE, Harris LW, Gold ER. Motivating participation in open science by examining researcher incentives. Elife 2017 Oct 30;6:e29319 [FREE Full text] [doi: 10.7554/eLife.29319] [Medline: 29082866]
4. Leonelli S, Spichtinger D, Prainsack B. Sticks AND Carrots: encouraging Open Science at its source. Geo 2015 Jun 30;2(1):12-16 [FREE Full text] [doi: 10.1002/geo2.2] [Medline: 26435842]
5. Sim I, Stebbins M, Bierer BE, Butte AJ, Drazen J, Dzau V, et al. Time for NIH to lead on data sharing. Science 2020 Mar 20;367(6484):1308-1309. [doi: 10.1126/science.aba4456] [Medline: 32193313]
6. Bezuidenhout L, Chakauya E. Hidden concerns of sharing research data by low/middle-income country scientists. Glob Bioeth 2018;29(1):39-54 [FREE Full text] [doi: 10.1080/11287462.2018.1441780] [Medline: 29503603]
7. Chawinga WD, Zinn S. Global perspectives of research data sharing: a systematic literature review. Library Inf Sci Res 2019 Apr;41(2):109-122. [doi: 10.1016/j.lisr.2019.04.004]
8. Fecher B, Friesike S, Hebing M, Linek S, Sauermann A. A reputation economy: results from an empirical survey on academic data sharing. ArXiv.org. URL: https://arxiv.org/abs/1503.00481 [accessed 2021-12-24]
9. Kim Y, Kim S. Institutional, motivational, and resource factors influencing health scientists' data-sharing behaviours. J Scholarly Publishing 2015 Jul;46(4):366-389. [doi: 10.3138/jsp.46.4.05]

10. Pisani E, Botchway S. Sharing individual patient and parasite-level data through the WorldWide Antimalarial Resistance Network platform: a qualitative case study. Wellcome Open Res 2017 Aug 16;2:63 [FREE Full text] [doi: 10.12688/wellcomeopenres.12259.1] [Medline: 29018840]
11. Cronin B. Hyperauthorship: a postmodern perversion or evidence of a structural shift in scholarly communication practices? J Am Soc Inf Sci 2001;52(7):558-569. [doi: 10.1002/asi.1097]
12. Global Research Report Multi-authorship and research analytics. Clarivate. URL: https://clarivate.com/webofsciencegroup/campaigns/global-research-report-multi-authorship-and-research-analysis/ [accessed 2021-12-22]
13. Larivière V, Desrochers N, Macaluso B, Mongeon P, Paul-Hus A, Sugimoto CR. Contributorship and division of labor in knowledge production. Soc Stud Sci 2016 Jun;46(3):417-435. [doi: 10.1177/0306312716650046] [Medline: 28948891]
14. Rennie D, Yank V, Emanuel L. When authorship fails. A proposal to make contributors accountable. JAMA 1997 Aug 20;278(7):579-585. [doi: 10.1001/jama.278.7.579] [Medline: 9268280]
15. Mongeon P, Larivière V. Costly collaborations: the impact of scientific fraud on co-authors' careers. J Assn Inf Sci Tec 2015 Jan 30;67(3):535-542. [doi: 10.1002/asi.23421]
16. Smith E, Williams-Jones B, Master Z, Larivière V, Sugimoto CR, Paul-Hus A, et al. Misconduct and misbehavior related to authorship disagreements in collaborative science. Sci Eng Ethics 2020 Aug;26(4):1967-1993 [FREE Full text] [doi: 10.1007/s11948-019-00112-4] [Medline: 31161378]
17. Gómez-Ferri J, González-Alcaide G, LLopis-Goig R. Measuring dissatisfaction with coauthorship: an empirical approach based on the researchers' perception. J Informetrics 2019 Nov;13(4):100980. [doi: 10.1016/j.joi.2019.100980]
18. Holcombe AO. Contributorship, not authorship: use CRediT to indicate who did what. Publications 2019 Jul 02;7(3):48. [doi: 10.3390/publications7030048]
19. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. N Engl J Med 2017 Apr 27;376(17):1684-1687. [doi: 10.1056/nejmsb1616595]
20. Wilsdon J, Bar-Ilan J, Frodeman R, Lex E, Peters I, Wouters P. Next-generation metrics: responsible metrics and evaluation for open science. Responsible Metrics for Higher Education & Research. URL: https://www.researchgate.net/publication/315385216_Next-generation_metrics_responsible_metrics_and_evaluation_for_open_science [accessed 2021-12-22]
21. Hicks D, Wouters P, Waltman L, de Rijcke S, Rafols I. Bibliometrics: The Leiden Manifesto for research metrics. Nature 2015 Apr 23;520(7548):429-431. [doi: 10.1038/520429a] [Medline: 25903611]
22. Declaration on Research Assessment (DORA). DORA Homepage. URL: https://sfdora.org/ [accessed 2021-12-22]
23. Altmetrics and rewards different types of Altmetrics - Thematic Report No 1. Mutual Learning Exercises on Open Science. URL: https://rio.jrc.ec.europa.eu/policy-support-facility/mle-open-science-altmetrics-and-rewards [accessed 2021-12-22]
24. Altmetrics and Rewards How to use altmetrics in the context of Open Science - Thematic Report No 2. Mutual Learning Exercises on Open Science. 2018. URL: https://rio.jrc.ec.europa.eu/policy-support-facility/mle-open-science-altmetrics-and-rewards [accessed 2021-12-22]
25. Recommendations of the OSPP on next-generation metrics. Altmetrics Working Group. URL: https://indico.egi.eu/event/5000/sessions/4493/attachments/13234/16015/integrated_advice_opspp_recommendations.pdf [accessed 2021-12-22]
26. Altmetrics and Rewards Implementing Open Science: Strategies, Experiences and Models - Thematic Report No 4. Mutual Learning Exercises on Open Science. URL: https://ec.europa.eu/research-and-innovation/en/statistics/policy-support-facility/mle-open-science-altmetrics-and-rewards [accessed 2021-12-22]
27. Altmetrics and Rewards Incentives and Rewards to engage in Open Science Activities - Thematic Report No 3. Mutual Learning Exercises on Open Science. URL: https://ec.europa.eu/research-and-innovation/en/statistics/policy-support-facility/mle-open-science-altmetrics-and-rewards [accessed 2021-12-22]
28. Biobanks in Europe: Prospects for Harmonisation and Networking. Luxembourg: Publications Office of the European Union; 2010.
29. Colledge FM, Elger BS, Shaw DM. "Conferring authorship": biobank stakeholders' experiences with publication credit in collaborative research. PLoS One 2013;8(9):e76686 [FREE Full text] [doi: 10.1371/journal.pone.0076686] [Medline: 24098803]
30. Pinel C. When more data means better results: abundance and scarcity in research collaborations in epigenetics. Soc Sci Inf (Paris) 2020 Mar 01;59(1):35-58 [FREE Full text] [doi: 10.1177/0539018419895456] [Medline: 32255899]
31. Pinel C. Renting valuable assets: knowledge and value production in academic science. Sci Technol Human Values 2021 Mar 01;46(2):275-297 [FREE Full text] [doi: 10.1177/0162243920911974] [Medline: 33518849]
32. Sauermann H, Haeussler C. Authorship and contribution disclosures. Sci Adv 2017 Nov;3(11):e1700404 [FREE Full text] [doi: 10.1126/sciadv.1700404] [Medline: 29152564]
33. Downe-Wamboldt B. Content analysis: method, applications, and issues. Health Care Women Int 1992;13(3):313-321. [doi: 10.1080/07399339209516006] [Medline: 1399871]
34. Graneheim UH, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today 2004 Feb;24(2):105-112. [doi: 10.1016/j.nedt.2003.10.001] [Medline: 14769454]
35. Schamber L. Time-line interviews and inductive content analysis: their effectiveness for exploring cognitive behaviors. J Am Soc Inf Sci 2000;51(8):734-744. [doi: 10.1002/(sici)1097-4571(2000)51:8<734::aid-asi60>3.0.co;2-3]

36. Defining the role of authors and contributors. International Committee of Medical Journal Editors. URL: http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html [accessed 2021-12-22]
37. Tornetta P, Siegel J, McKay P, Bhandari M. Authorship and ethical considerations in the conduct of observational studies. J Bone Joint Surg Am 2009 May;91 Suppl 3:61-67. [doi: 10.2106/JBJS.H.01538] [Medline: 19411501]
38. Birnholtz JP. What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. J Am Soc Inf Sci 2006 Nov;57(13):1758-1770. [doi: 10.1002/asi.20380]
39. Murtagh MJ, Turner A, Minion JT, Fay M, Burton PR. International data sharing in practice: new technologies meet old governance. Biopreserv Biobank 2016 Jun;14(3):231-240. [doi: 10.1089/bio.2016.0002] [Medline: 27200470]
40. Rice DB, Raffoul H, Ioannidis JP, Moher D. Academic criteria for promotion and tenure in biomedical sciences faculties: cross sectional analysis of international sample of universities. BMJ 2020 Jun 25;369:m2081 [FREE Full text] [doi: 10.1136/bmj.m2081] [Medline: 32586791]
41. Helgesson G, Eriksson S. Authorship order. Learned Publishing 2018 Aug 21;32(2):106-112. [doi: 10.1002/leap.1191]
42. Mongeon P, Smith E, Joyal B, Larivière V. The rise of the middle author: investigating collaboration and division of labor in biomedical research using partial alphabetical authorship. PLoS One 2017;12(9):e0184601 [FREE Full text] [doi: 10.1371/journal.pone.0184601] [Medline: 28910344]
43. Smith E, Master Z. Best practice to order authors in multi/interdisciplinary health sciences research publications. Account Res 2017;24(4):243-267. [doi: 10.1080/08989621.2017.1287567] [Medline: 28128975]
44. Mazumdar M, Messinger S, Finkelstein DM, Goldberg JD, Lindsell CJ, Morton SC, Biostatistics, Epidemiology,Research Design (BERD) Key Function Committee of the ClinicalTranslational Science Awards (CTSA) Consortium. Evaluating academic scientists collaborating in team-based research: a proposed framework. Acad Med 2015 Oct;90(10):1302-1308 [FREE Full text] [doi: 10.1097/ACM.0000000000000759] [Medline: 25993282]
45. Smith E, Williams-Jones B, Master Z, Larivière V, Sugimoto CR, Paul-Hus A, et al. Researchers' perceptions of ethical authorship distribution in collaborative research teams. Sci Eng Ethics 2020 Aug;26(4):1995-2022 [FREE Full text] [doi: 10.1007/s11948-019-00113-3] [Medline: 31165383]
46. Pinel C, Prainsack B, McKevitt C. Caring for data: value creation in a data-intensive research laboratory. Soc Stud Sci 2020 Apr 13;50(2):175-197 [FREE Full text] [doi: 10.1177/0306312720906567] [Medline: 32053062]
47. Soares MB. Collaborative research in light of the prevailing criteria for promotion and tenure in academia. Genomics 2015 Oct;106(4):193-195 [FREE Full text] [doi: 10.1016/j.ygeno.2015.07.009] [Medline: 26232606]
48. Vasilevsky NA, Hosseini M, Teplitzky S, Ilik V, Mohammadi E, Schneider J, et al. Is authorship sufficient for today's collaborative research? A call for contributor roles. Account Res 2021 Jan;28(1):23-43. [doi: 10.1080/08989621.2020.1779591] [Medline: 32602379]
49. Brand A, Allen L, Altman M, Hlava M, Scott J. Beyond authorship: attribution, contribution, collaboration, and credit. Learn Pub 2015 Apr 01;28(2):151-155. [doi: 10.1087/20150211]
50. Allen L, O'Connell A, Kiermer V. How can we ensure visibility and diversity in research contributions? How the Contributor Role Taxonomy (CRediT) is helping the shift from authorship to contributorship. Learned Publishing 2019 Jan 24;32(1):71-74. [doi: 10.1002/leap.1210]
51. Frische S. It is time for full disclosure of author contributions. Nature 2012 Sep 27;489(7417):475. [doi: 10.1038/489475a] [Medline: 23018928]
52. McNutt MK, Bradford M, Drazen JM, Hanson B, Howard B, Jamieson KH, et al. Transparency in authors' contributions and responsibilities to promote integrity in scientific publication. Proc Natl Acad Sci U S A 2018 Mar 13;115(11):2557-2560 [FREE Full text] [doi: 10.1073/pnas.1715374115] [Medline: 29487213]
53. No authors listed. Uniform requirements for manuscripts submitted to biomedical journals. International Committee of Medical Journal Editors. JAMA 1997 Mar 19;277(11):927-934. [Medline: 9062335]
54. Huth E, Case K. The URM: twenty-five years old. Sci Editor 2004 Feb;27(1):17 [FREE Full text]
55. Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals. International Committee of Medical Journal Editors. 2013. URL: http://www.icmje.org/recommendations/archives/ [accessed 2021-12-22]
56. Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals. International Committee of Medical Journal Editors. 2017. URL: http://www.icmje.org/recommendations/archives/ [accessed 2021-12-22]
57. Samet JM. Data: to share or not to share. Epidemiology 2009 Mar;20(2):172-174. [doi: 10.1097/ede.0b013e3181930df3]
58. Khoury MJ. Planning for the future of epidemiology in the era of big data and precision medicine. Am J Epidemiol 2015 Dec 15;182(12):977-979 [FREE Full text] [doi: 10.1093/aje/kwv228] [Medline: 26628513]
59. Ness RB. "Big" science and the little guy. Epidemiology 2007 Jan;18(1):9-12. [doi: 10.1097/01.ede.0000249529.86885.5a] [Medline: 17179753]
60. Establishing incentives and changing cultures to support data access. Expert Advisory Group on Data Access. URL: https://wellcome.org/sites/default/files/establishing-incentives-and-changing-cultures-to-support-data-access-eagda-may14.pdf [accessed 2021-12-22]

Abbreviations

CRediT: contributor roles taxonomy
ICMJE: International Committee for Medical Journal Editors

Edited by G Eysenbach; submitted 24.11.20; peer-reviewed by B Prainsack, V Meschesi; comments to author 08.03.21; revised version received 14.03.21; accepted 19.11.21; published 13.01.22.

Please cite as:
Devriendt T, Borry P, Shabani M
Credit and Recognition for Contributions to Data-Sharing Platforms Among Cohort Holders and Platform Developers in Europe: Interview Study
J Med Internet Res 2022;24(1):e25983
URL: https://www.jmir.org/2022/1/e25983
doi:10.2196/25983
PMID:35023849

Original Paper

Ethical Issues of Digital Twins for Personalized Health Care Service: Preliminary Mapping Study

Pei-hua Huang[1*], PhD; Ki-hun Kim[2*], PhD; Maartje Schermer[1], PhD

[1]Department of Medical Ethics, Philosophy and History of Medicine, Erasmus MC University Medical Center, Rotterdam, Netherlands

[2]Department of Industrial Engineering, Pusan National University, Busan, Republic of Korea

[*]these authors contributed equally

Corresponding Author:
Pei-hua Huang, PhD
Department of Medical Ethics, Philosophy and History of Medicine
Erasmus MC University Medical Center
PO Box 2040
Rotterdam, 3000CA
Netherlands
Phone: 31 650031752
Email: p.huang.1@erasmusmc.nl

Abstract

Background: The concept of digital twins has great potential for transforming the existing health care system by making it more personalized. As a convergence of health care, artificial intelligence, and information and communication technologies, personalized health care services that are developed under the concept of digital twins raise a myriad of ethical issues. Although some of the ethical issues are known to researchers working on digital health and personalized medicine, currently, there is no comprehensive review that maps the major ethical risks of digital twins for personalized health care services.

Objective: This study aims to fill the research gap by identifying the major ethical risks of digital twins for personalized health care services. We first propose a working definition for digital twins for personalized health care services to facilitate future discussions on the ethical issues related to these emerging digital health services. We then develop a process-oriented ethical map to identify the major ethical risks in each of the different data processing phases.

Methods: We resorted to the literature on eHealth, personalized medicine, precision medicine, and information engineering to identify potential issues and developed a *process-oriented ethical map* to structure the inquiry in a more systematic way. The *ethical map* allows us to see how each of the major ethical concerns emerges during the process of transforming raw data into valuable information. Developers of a digital twin for personalized health care service may use this map to identify ethical risks during the development stage in a more systematic way and can proactively address them.

Results: This paper provides a working definition of digital twins for personalized health care services by identifying 3 features that distinguish the new application from other eHealth services. On the basis of the working definition, this paper further layouts 10 major operational problems and the corresponding ethical risks.

Conclusions: It is challenging to address all the major ethical risks that a digital twin for a personalized health care service might encounter proactively without a conceptual map at hand. The process-oriented ethical map we propose here can assist the developers of digital twins for personalized health care services in analyzing ethical risks in a more systematic manner.

(J Med Internet Res 2022;24(1):e33081) doi:10.2196/33081

KEYWORDS

digital twins; digital health; personalized health care service; data-driven health care; value-sensitive design; ethics of health care technology

Introduction

Background

The concept of digital twins is expected to transform the landscape of existing health care systems [1,2]. Originating from industrial design, the concept of digital twins capitalizes on the data of specific objects to simulate replicas in the virtual world for predictive analysis of safety risks and testing of different optimization solutions. Applying to the health care sector, a digital twin can be a virtual replica of a particular patient that reflects the unique genetic makeup of the patient [3] or a simulated 3D model that exhibits the characteristics of a patient's heart [4,5]. Ideally, these digital twins will allow clinicians to offer personalized health care to individual patients.

Digital twins may make health care services more proactive and personalized. With predictive algorithms and real-time data, digital twins have the potential to detect anomalies and assess health risks before a disease develops or becomes symptomatic. Information provided by digital twins can then help clinicians determine whether early intervention is necessary [6,7]. Digital twins equipped with data such as patients' genetic information and disease history may also facilitate clinicians to *personalize* the treatment [8,9]. Each individual has a unique genetic makeup, experiences different diseases, and lives in a different environment. These differences also make individual patients respond to different treatments in different ways. Under certain circumstances, the response could be so adverse that the treatment meant to heal the patient causes hospitalization or even death. It is envisioned that in the future, clinicians may simulate the effects of all possible treatments on patients' digital twins first and determine which option is likely to be the most effective for the patients.

The idea of digital twins for personalized health care is gaining traction from both the public and private sectors. Several research consortia have been established to explore the potential of digital twins for personalized health care services. In the Netherlands, the Erasmus MC University Medical Center, Delft University of Technology, and Erasmus University Rotterdam have jointly initiated research on digital twins and cardiovascular disease prevention [10]. Sweden's Linköping University also launched a pioneering research project, MeDigiT [11], to explore the potential of digital twins for medical education, heart disease diagnosis, medical implant planning, and so forth. Another 2 major research groups in Europe, the Swedish Digital Twin Consortium [12] and the DigiTwins [13], focus on translating the concept of digital twins into fields such as molecular medicine and genomic research. In the private sector, Philips has rolled out a clinical application called HeartModel that allows cardiologists to plan their upcoming surgeries with high-resolution interactive 3D models that reflect the distinctive and unique features of their patients' hearts [4]. Siemens Healthineers and GE Healthcare also have similar products under development [5].

However, the growing interest in applying the concept of digital twins to personalized health care has raised new questions. Indeed, the concept of a digital twin is well-established and widely applied in industrial design and engineering. There is consensus about what sort of data and infrastructure are required to develop a digital twin for purposes such as predictive maintenance and optimization planning [7]. Nonetheless, the *objects* involved in the health care sector are very different from those in the realm of engineering. Even with the completion of the Human Genome Project, our understanding of genes and diseases is still very limited. Developing a digital twin that can adequately simulate or predict a person's health condition is much more challenging than building a digital replica of a nonliving object. In addition, humans are living beings with personal commitments and moral worldviews. Applying the engineering concept to health care without taking these differences into account will be deeply problematic.

To our knowledge, there is currently no comprehensive review on the ethical risks of developing personalized health care services based on the concept of digital twins. Indeed, some have already noted that the data-driven nature of digital twins requires developers to pay special care to privacy protection [14]. However, privacy is only one of the ethical risks developers of digital twins need to carefully address during development. The involvement of predictive algorithms, for instance, could expose users to algorithmic biases [15]. Bruynseels et al [16] also argue that digital twins might worsen existing inequalities. Those who are less well-off might not have the means to take advantage of the service because of a lack of access to devices that can precollect the data required for using the service. Given the ever-increasing interest in transforming the health care sector with the concept of digital twins, it is urgent to identify the major ethical risks of digital twins for personalized health care services.

Objectives

This research aims to address this research gap by providing a comprehensive analysis of the major ethical risks of digital twins for personalized health care services. Owing to the scarcity of literature on digital twins for personalized health care services, we are unable to perform a systematic review of ethical concerns over digital twins for personalized health care services. As a convergence of health care, artificial intelligence, information and communication technologies (ICTs), and personalized health care services, it is also difficult to apply the existing bioethical framework to capture the distinctive features and corresponding ethical risks of digital twins for personalized health care services. In addition, as influential bioethical frameworks such as the Beauchamp and Childress [17] 4 principles framework focus on high-level abstract ethical principles, it is challenging to translate the principles into specific and concrete normative guidance for first-line developers.

Methods

Owing to the challenges stated in the *Introduction* section, we resorted to the literature on eHealth, personalized medicine, precision medicine, and information engineering to identify potential issues. We developed a *process-oriented ethical map* to structure the inquiry in a more systematic way. The *ethical map* allows us to see how each of the major ethical concerns emerges during the process of transforming raw data into

valuable information. Developers of a digital twin for personalized health care service may use this map to identify ethical risks during the development stage in a more systematic way and proactively address them.

We first provided a working definition of digital twins for personalized health care services to clarify the extent to which our research was applicable. We then consulted the literature on ICTs to develop a process-oriented ethical map to structure the identified ethical risks. We have discussed the limitations of our research and provided recommendations at the end of this paper.

Results

Digital Twins for Personalized Health Care Service: Working Definition

Overview

Despite all the interest in translating the concept of digital twins to personalized health care services, there is *no* consensus on the definition of digital twins for personalized health care services. Some services focus on visualization, whereas others aim to offer predictive analysis. Some capitalize on existing data, whereas others require continuous input. Differences such as these create challenges in providing a systematic analysis of potential ethical pitfalls of digital twins for personalized health care service.

The lack of consensus is attributable to 2 reasons. The first reason is the ambiguity of *personalization* [15]. Each patient is different in various aspects. Every individual has different molecular and genomic features, and no patient has the same socioeconomic background, preferences, needs, and conception of the good. Therefore, personalized health care could be understood in at least two ways. When focusing on the biological aspect, the meaning of personalization is akin to precision. Following this interpretation of personalization, the general goal of personalized health care is to fine-tune health care with health-related data and administer treatments that are likely to be most effective and cause the fewest adverse side effects to a patient. Achieving personalization in this sense does not require active participation from the patient. In contrast, when focusing on the nonbiological aspect, to realize personalization is to respect individual patients' personal commitments and values. Personalized health care, in this sense, is a health care ideal that aims to give back agency to the patient, facilitating the patient to autonomously choose the treatment course that can best reflect their values or cater to their particular needs. Thus, the improvement of the patient's physical health would not be the primary consideration here.

The second reason is the difference in goals. Each personalized health care service aims to address different health issues. The solutions adopted by developers might also vary. Phillip's HeartModel provides a personalized health care service in the sense that the digital twin reflects the unique anatomical structure of a particular patient. Their goal was to improve the quality of the surgery. As for My Digital Twin, developed by the Dutch research team, the goal was to use a digital twin to crunch health-related data to predict whether a person was on the trajectory of developing cardiovascular diseases. Although the 2 cases shared a general goal, that is, improving health care quality, the health issues they aimed to address and the services they aimed to provide were very different.

In this study, we do not aim to provide a definitive account of what a digital twin for personalized health care services ought to be. However, to begin a critical analysis of potential ethical pitfalls, we propose a working definition to clarify the extent to which our ethical framework is applicable.

Definition of a Digital Twin for Personalized Health Care Service

A digital twin for personalized health care service is a *data-driven, interactive computerized model* that aims to offer *health-related information* that properly simulates or predicts the health conditions of a *particular person*.

Data-Driven, Interactive Computerized Model

Any attempt to incorporate the concept of digital twins in a health care system requires an input of data. The general idea behind the concept of digital twins is to capitalize on precollected or real-time data to build up interactive models that allow users to conduct various simulations (eg, descriptive modeling, predictive analysis of risk levels, or prescriptive recommendation). In the health care context, the general goal of a digital twin for personalized health care service is to capitalize on data that are directly or indirectly related to an individual patient's health conditions to build up computerized models that allow users (the patient or relevant clinicians) to gain an opportunity to devise and test different virtual trials (eg, lifestyles, pharmaceutical interventions, and surgical approaches) on the patient's digital twin. The data used by a digital twin for personalized health care service can be identifiable data or nonidentifiable data. For instance, a digital twin's predictive algorithms can be trained with multiple deidentified data sets and can make predictions on a person's health trajectory based on certain identifiable data provided by the person.

Health-Related Information

Data treated properly can yield 3 types of health-related information [18]. *Descriptive information* indicates what has happened or is happening to a person's health. *Predictive information* offers insights regarding what is likely to happen to a person's health. *Prescriptive information* provides suggestions regarding which action or intervention should be adopted for the sake of improving or restoring a person's health. These 3 information types are essential for personalized health monitoring, diagnosis, prognosis, prevention, and treatment.

Depending on the goals, the information provided by a digital twin for personalized health care service may involve only 1 type of information or multiple types of information. A simple 3D model of a patient's heart, for instance, might only deliver descriptive information. In contrast, a model built from a patient's genomic data might offer more than a mere description of the patient's health. For instance, a digital twin built from a person's genomic data has the potential to predict the effectiveness of a particular treatment course for a specific

patient and prescribe treatment recommendations for a specific patient.

Particular Person

The computerized model or prediction generated by a digital twin must properly simulate the unique characteristics of a person. In this sense, incorporating the concept of a digital twin into the health care sector may yield a new form of personalized health care service.

However, the focus on simulating the health conditions of a particular person does not mean that a digital twin for personalized health care services can only use data from the person. Calibration of the algorithms used for simulation or prediction may require a large amount of health-related data collected from the general public. For instance, providing more personalized advice regarding blood pressure and hypertension management will require the developers to first work with relevant data to refine the baseline blood pressure.

This working definition helps us differentiate a digital twin for personalized health care service from general digital health care (or eHealth). For instance, although telehealth also capitalizes on ICTs and arguably requires data input from the patient (eg, via teleconsultation), telehealth services do not depend on computerized modeling. Instead, the value of telehealth results mainly from offering patients the opportunity to consult their clinicians remotely. The requirement of interactability also helps us distance a digital twin for personalized health care service from medical technologies that have long been adopted to create digital images of particular persons. For instance, although magnetic resonance imaging (MRI) also relies heavily on ICTs to transform the collected data into imagery information of specific persons, these computerized images do not offer clinicians the opportunities to conduct virtual trials on them. As a result, MRI does not qualify as a digital twin for personalized health care service but is a medical device that may be incorporated into a digital twin for personalized health care service. One of Linköping University's MeDigiT projects, for instance, uses MRI and computed tomography data to simulate a heart digital twin to better personalize the artificial heart implant (Table 1).

Table 1. Examples of different types of digital twins for personalized health care service.

Type of digital twin	Type of primary data used by the digital twin	Forms of digital twin	Types of information delivered
HeartModel [4]	Imagery data (ultrasound)	Interactive visual presentation of the anatomical and physiological features of the heart	Descriptive
MeDigiT [11]	Imagery data (MRI[a] and CT[b])	Interactive visual presentation of the anatomical and physiological features of a specific part of the body	Descriptive
My Digital Twin [10]	Lifestyle data (dietary, smoking, use of alcohol, and medication), environmental data (living and working situations), and electronic health records (visits to health care services, medication, biotest results, MRI scans, and CT scans)	An aggregated model that offers information about a person's current health conditions, a prediction of relevant health risks, and health advice on improving the health condition	Descriptive, predictive, and prescriptive
Personalized diagnosis and therapy via genomic medicine [3]	Genomic data	A genomic model allows users to identify treatments that are likely to be most effective for a particular patient	Predictive and prescriptive

[a]MRI: magnetic resonance imaging.

[b]CT: computed tomography.

A Process-Oriented Ethical Map

Overview

Digital twins for personalized health care services aim to synthesize valuable information from health-related data for timely diagnosis, prognosis, preventive intervention, or treatment optimization. Achieving these goals requires a sophisticated orchestration of multiple ICTs and the involvement of various stakeholders in tasks such as data collection, data analysis, and information presentation. Each of the data processing phases faces different ethical risks. It is challenging to address all the major ethical risks that a digital twin for personalized health care service might encounter proactively without a conceptual map at hand. The process-oriented ethical map we have proposed below would assist developers of digital twins for personalized health care services in analyzing ethical risks in a more systematic manner (Table 2).

Briefly, despite the complex infrastructure, the process of creating valuable information can be divided into four major phases: data collection, data management, data analysis, and information use. Each of the 4 major phases requires different ICTs and information systems to realize its desired goal. It is not surprising that each of the 4 phases touches on various ethical issues. For instance, to continuously collect data from the user, developers of a digital twin for personalized health care services would need to address worries about surveillance and issues related to data accessibility. However, even if the developers properly deal with these issues, ill-designed algorithms might still cause great harm to users of the digital twin for personalized health care services by offering them a distorted picture of a person's health conditions. In a situation where the digital twin for personalized health care services is free from data-related concerns such as data collection and data analysis, the service could still induce negative influences on

its users by taking an overly demanding concept of health as the norm (coercive healthism).

This process-oriented framework shows that although many ethical concerns are interrelated, they might arise independently in different phases. Breaking down a digital twin for personalized health care service into 4 major phases helps developers of this service conduct an ethical assessment during the design stage in a more systematic way.

Table 2. A process-oriented ethical map.

Operation process and operational problem	Ethical issues
Data collection	
Hypercollection	• Autonomy • Informed consent • Right to privacy • Surveillance health care
Data quality and unorthodox use	• Distortion of the understanding of health
Data management	
Data ownership and data accessibility	• Autonomy • Health equity
Data ownership and data brokerage	• Autonomy or informed consent • Right to privacy • Transparency
Hacking	• Right to privacy
Data analysis	
Biased algorithms	• Discrimination or injustice • Distortion of the understanding of health
Biased training data set	• Discrimination or injustice • Distortion of the understanding of health
Information use	
Decontextualization of disease formation	• Autonomy • Distortion of the understanding of health • Victim blaming
Epistemic injustice	• Autonomy • Distortion of the understanding of health • Damage physician–patient relationship
Overdiagnosis	• Distortion of the understanding of health • Right to bodily integrity

Data Collection

Overview

Data collection is an indispensable phase of any digital twin for personalized health care service. All data analyses and simulations require initial data input. However, the potential for gaining more information about a person also exposes the person to several ethical risks. On the normative side, practices such as hypercollection can severely infringe on the *right to privacy* and *autonomy*. With no clear understanding of the scope of data collection, meaningful *informed consent* is often missing. As a service that aims to provide better personalized health care, digital twins for personalized health care services also face several ethical risks from the epistemic side. The quality of the collected data might not be good enough to achieve the desired goals, such as providing a more comprehensive understanding of a person's health conditions or making an accurate prediction of a person's likelihood of developing certain diseases.

Hypercollection

To construct proper models for personalized analysis, a digital twin for personalized health care service might need to access various data sets to train and recalibrate the algorithms used for data analysis. From an engineering perspective, health-related data can be defined as any data that can contribute to drawing inferences on a person's health condition. It might be tempting to incorporate data about one's social media use, education, occupation, and other sources that are not traditionally viewed as health-related data in a digital twin [19]. However, there is also a growing concern that service providers might secretly exploit the data collection process by collecting as much data

as possible, although some of the collected data are not relevant to the service the digital twin for personalized health care service aims to provide [15].

Furthermore, even if a digital twin for personalized health care service only requests data that fall under the traditional understanding of health data (eg, electronic health records and biopsy), the developer still has to justify the necessity of including the requested health data. Health data are widely considered highly sensitive. Physicians and health care organizations as patients' fiduciaries have special moral duties to promote patients' well-being and protect patients' privacy [17]. Legally speaking, health data are also subject to stringent legal protection [20]. To access a particular set of health data, one must provide a strong reason to justify why the set of health data is necessary for the task and proactively request informed consent from the patients.

In addition, requesting extensive data from the users might also put the users under undue risks of inference attack, a data mining technique that uses authorized data to access authorized information via inference and common knowledge [21]. The requested data can be used to reveal information that users do not wish to share with the developers, seriously infringing on the users' right to privacy. Merely stating that inclusion may enhance the predictive power and accuracy of a digital twin for personalized health care services is not sufficient to outweigh the privacy concerns.

The growing accessibility of wearables and biosensors offers developers of digital twin systems opportunities to build up a system that can update a person's digital twin in real time. The pharmaceutical company, Otsuka, has developed a new generation of digital pills (Abilify MyCite) that helps patients track medicine intake by sending signals to the patient's mobile devices and relevant parties [22,23]. These technologies are usually marketed as innovations that can empower patients by helping them better manage their health conditions (eg, improving adherence). However, many bioethicists cast doubt on this rhetoric. Some physicians might cajole or coerce their patients to take the digital pills so that they can *monitor* their patients [24,25]. Despite being informed, patients might not truly consent to be monitored by taking ingestibles. Circumstances such as this could increase patients' anxiety levels and reduce trust between physicians and their patients [26].

Data Quality and Unorthodox Use

Another issue related to data collection is data quality and accuracy. Indeed, wearables now make the collection of a wide range of biosignals possible. However, the accuracy of the devices used for data collection varies. Consider the Apple Watch as an example. Despite the increasing interest in incorporating this device into the digital twin ecosystem, a recent review on the accuracy of the Apple Watch's performance in measuring heart rate and energy expenditure found that although the device offers clinically reliable measurement of heart rates, it systematically overestimates the expenditure of energy in patients with cardiovascular disease [27]. Marcus [28] also pointed out that the false-positive rates were unacceptably high in an Apple-sponsored research on atrial fibrillation. Only 35% of the research participants who participated in the validation phase (n=450) presented with atrial fibrillation when examined by a traditional electrocardiogram.

These studies show that although non–medical-level wearables offer an affordable way for the general public to trace and manage their lifestyle, the accuracy of the data gathered by these devices does not always meet the clinical standards. Instead of paving the path for a more personalized health care service, attempts to capitalize data collected from commercial-level wearables might risk creating a distorted digital image of people. Developers must carefully consider the level of data accuracy required for the services they are developing.

The reliability of a digital twin is also vulnerable to unorthodox use of the service. A user of the digital twin might not follow the instructions properly and therefore compromise the quality or accuracy of the data collected by the device. Some users might deliberately use the device in an unorthodox way to *trick the system* in certain circumstances. For instance, a digital twin for personalized health care service devised by insurance companies could be compromised as some users might be more interested in getting a lower premium rather than tracking how a newly adopted healthy lifestyle could improve their health with the digital twin [29]. If a compromised digital twin is to be linked to other general medical services, the compromised digital twins might also undermine a clinician's capability to make sound clinical judgment. Developers must take precautionary steps to minimize such risks.

Data Management

Overview

Developers may devise very different management strategies to optimize their services. However, the differences in management strategies can also create obstacles to data accessibility, diminishing users' *autonomy* in terms of seeking the best use of their data as they see fit. Certain providers might also engage with data brokerage and sell the entrusted data for profits. Although data brokerage is not inherently unethical, selling sensitive health-related data without explicit consent fails to show due respect for the *right to privacy*. The complicated ICT ecosystem of digital twins for personalized health care services might also expose users to undue hacking risks.

Data Accessibility

Digital obsolescence may affect people's ability to reuse their data for other health care services should the service provider fail to devise proper management strategies after each system upgrade [30]. In addition, it is foreseeable that some of the developers of digital twins for personalized health care services would face a close-down and cease to offer service maintenance thereafter. The disruption of service might create difficulties for users of the digital twins for personalized health care services to retrieve the health-related data they entrusted to the service providers. The fail-fast culture of technology startups might exacerbate this problem. Given that digitalization of health care is an unstoppable trend, the inability to access one's data would severely affect the quality of health care a person can receive. It is important to recognize these accessibility issues and devise

means that allow the users of digital twins for personalized health care services to access, retrieve, and transfer the data they have entrusted to their service provider.

Data Brokerage

Despite a lack of consensus on how to characterize data ownership and whether the right to data ownership exists [31], data brokerage as a business model is prevalent in the mobile health industry [32]. Health-related data such as patient experience, medical history, and symptoms are especially valuable to pharmaceutical companies and marketing organizations as they may improve drug development and marketing strategies [15,33]. Although data brokerage as a business model is not inherently unethical, many service providers fail to obtain explicit informed consent from their users. Huckvale et al [34] recently found that of the 36 top-ranked Android and iOS apps for depression and smoking cessation, 29 apps transmitted the entrusted data to Facebook or Google (sometimes both) for advertising and analytics services. Only 12 apps accurately disclosed this practice. Given the levels of sensitivity of health-related data, selling them without obtaining explicit consent from the users might severely affect the users' right to privacy. Service providers ought to convey their plans, if any, for the secondary use of the entrusted data to relevant parties transparently and seek explicit informed consent from the users.

Hacking

The digitalization of health care has also attracted the attention of malicious hackers [35]. A survey conducted by KPMG [36] also showed that 81% of the 223 surveyed organizations experienced cyberattacks. In another study conducted by the Institute for Critical Infrastructure Technology [37], it was estimated that >110 million patients in the United States had their health data compromised in 2015 alone. Given that the promise of digital twins for personalized health care services is built on extensive health-related data, they might attract even more cyberattacks than other services in the health care sector have ever undergone. Developers of digital twins for personalized health care services must invest in cybersecurity to properly safeguard the data entrusted to them and the operation of the systems.

Data Analysis

Overview

Data processing is an essential phase for extracting and synthesizing information from otherwise fragmented and uninformative data. Well-designed algorithms can reveal valuable information that can enhance decision-making capacity. However, this power also makes algorithms become a double-edged sword—they can be used to crunch accessible data to reveal unauthorized information, posing a great threat to people's right to privacy. The human tendency to trust automatic systems may make users of a digital twin susceptible to harm brought about by biased algorithms.

Biased Algorithms

Algorithms are the backbone of any data-driven health care service. They execute the instruction designed by human developers, sort and weigh various data, and produce the desired information such as risk assessment and prognosis. Although algorithms are not liable to influences such as emotions and fatigue, they could still yield unanticipated discriminatory results. Obermeyer et al [38] recently discovered that Black patients were systematically discriminated against by a widely adopted health care algorithm for identifying patients who are highly likely to need complex health care. The algorithm unintentionally discriminated against Black patients by assigning them lower risks as it used health care costs as a proxy for prediction. It is generally true that the more complex the health needs, the higher the cost. However, using health care costs as a proxy overlooks the fact that expenditure depends partially on health care access. The lower amount of health expenditure observed in Black patients does not imply that they are less ill than White patients. Instead, it is more likely to result from unequal access to health care. Obermeyer et al [38] also found that once replaced by the inappropriate proxy used by the system, patients with African backgrounds could have received additional support from 17.7% to 46.5%. This study shows that developers of a digital twin for personalized health care services must pay extra attention to calibrating and validating the algorithms used in the system.

Biased Training Data Set

A digital twin for personalized health care services might incorporate advanced computing technologies such as deep learning and machine learning in the data analysis phase. The powerful technologies can be used to detect hidden correlations between different variables, assisting the digital twin in predictive analysis for health risk assessment or treatment outcome assessment. However, the reliability of deep learning and machine learning can be severely compromised if the data sets used to train these algorithms do not properly reflect the environment in which these algorithms are to navigate [39]. Liu et al [40] found that IBM's Watson for Oncology was less effective and reliable when applied to non-Western populations as the imagery data used for training Watson were primarily from the Western population. Recently, it was also found that certain data sets used for training machine learning algorithms are, in fact, unfit for the task. The labeling of the chest X-ray images in the ChestXray14 database were not standardized and sometimes did not match with the image at all [41]. Similar problems were also identified in machine learning research that aimed to capitalize on chest radiographs and computed tomography scans to detect COVID-19 [42]. Developers must ensure that the training data reflects the characteristics of the served population and are correctly and consistently labeled. Otherwise, the predictive analysis offered by the digital twin can be misleading and even discriminatory, bringing more harm than benefit to the users.

Information Use

Overview

The use of health-related information is not a value-free practice. The decision regarding which information is worth presenting conveys the values upheld by the developers of a digital twin for personalized health care services. In the context of predictive analysis, the risk scores a digital twin gives to an individual

reflect the conception of health and disease for the developers of the digital twin. Without careful reflection, the developers of the digital twin risk passing down problematic values such as victim-blaming culture and distrust of personal experiences. The goal of earlier diagnosis and intervention could also lead to overdiagnosis and infringement of people's bodily integrity.

Decontextualization of Disease Formation

A digital twin for personalized health care services might overly individualize health issues and overlook the fact that socioenvironmental determinants, such as air pollution, water pollution, and lack of education, also contribute to health problems [43,44]. Victims of environmental pollution and social injustice might be wrongfully blamed for their poor health. In addition, although a digital twin for personalized health care service may allow people to access health information they otherwise could not access, the epistemic improvement does not warrant empowerment. People with lower socioeconomic backgrounds might not know how to use the provided information or not have the agency to act upon the information because of external constraints [45]. Contrary to the goal of empowerment, the digital twin for personalized health care services might burden patients with a sense of powerlessness, guilt, and anxiety. It is especially so for a digital twin for personalized health care service that aims to introduce early interventions in lifestyle diseases such as diabetes, hypertension, and obesity [46]. Users who fail to take the advised change could be accused of being irresponsible about their health (victim blaming).

Epistemic Injustice

The growing reliance on health information produced by digital twins for personalized health care services could also lead to undervaluing patients' personal views and experiential knowledge. Some might think that health information offered by the digital twin is more reliable than a patient's personal account as the information results from an *objective fact*. However, this view overlooks the fact that this information was generated from a system developed with a human's limited understanding of human biology and other relevant fields. The information would be full of human interpretation and subject to various biases as well. Rich et al [47] recently found that the discrepancies between the analysis by fitness apps and users' subjective feelings support the concern over epistemic injustice. Downplaying the patient's experiential knowledge simply because this piece of knowledge has subjective elements is deeply problematic. Instead of offering a more holistic understanding of health, the digitalization of health could create a distorted understanding of health [15,48].

Overdiagnosis

Another concern related to the definition of health is overdiagnosis. One of the general goals of digital twins for personalized health care services is to provide early warnings to its users and assist in preventive health care. However, in practice, early action sometimes leads to overdiagnosis and overtreatment. This sort of ethical dilemma has been highlighted in the personalized medicine literature on the use of biomarkers [49-51]. For example, many bioethicists and clinicians are concerned that genetic testing that can be used to detect *BRCA1* and *BRCA2* mutations might cause overtreatment [52,53], causing harm to a patient's bodily integrity.

Furthermore, when gaining access to more information about various health-related parameters, it is important to reflect on the extent to which deviations from *the norm* can be considered diseases. Sexuality is a prominent example of this. Hormone levels can also differ significantly between women with and without pregnancies [54]. The conceptual link to nonbinary concepts such as dysfunction, harm, and risk also suggests that there is no clear line to be drawn between diseased and nondiseased states. Following this observation, Walker and Rogers [49] argue that *overdiagnosed cases* can be understood as *borderline cases* that are neither diseased nor healthy but in between. Treating borderline cases and those that are clearly diseased in the same way is morally problematic. Recent advocacy of renaming low-risk conditions that are unlikely to develop into cancers echoes this concern [55]. There is also a growing number of bioethicists and medical practitioners casting doubt on the utility of detecting borderline cases [50]. Without careful stratification and selection of reference groups, a digital twin for personalized health care services might risk providing wrongful health advice to its user. Therefore, developers of a digital twin for personalized health care services ought to consult clinical practitioners and relevant researchers to fine-tune the system with comprehensive epidemiological knowledge.

Discussion

Limitations of the Study

The analysis we performed in this research offers a clear overview of the major operational problems that might damage vital ethical values during each of the data processing and information use stages. However, this ethical analysis of digital twins for personalized health care services has several limitations. First, as digital twins for personalized health care services are still in their infancy, the literature directly addressing digital twins for personalized health care services is scarce. Most of the ethical analyses we conducted here is based on the literature in fields that we considered closely linked to digital twins for personalized health care services. However, mapping the major ethical risks in this way renders the analysis heavily influenced by our prior knowledge, and we might have overlooked certain ethical risks of digital twins for personalized health care services. Second, as the process-oriented framework we proposed in this research aims to provide a conceptual map to help developers proactively examine potential ethical risks that might occur in each of the major data processing phases, the framework would be less effective in facilitating developers to examine ethical risks based on the type of information provided by a digital twin for the health care system. For instance, it is a known ethical risk that genetic information can be used to infer the health conditions of a person's family members. For people who do not want to know whether they are at risk of certain genetic diseases, their right not to know might be infringed by the family member who decided to use the information [56,57]. However, because of the structure of the framework, there is no room for this important discussion. It is desirable to see further research on health information

generated by a digital twin for personalized health care services and the associated ethical risks. Third, there could be novel ethical risks of digital twins for personalized health care services that are distinctively different from the concerns that have been identified in the literature on digital health, personalized medicine, and precision medicine. This paper is by no mean trying to provide a definitive account of the ethicality of digital twins for personalized health care services. Further empirical ethics research on digital twins for personalized health care services is necessary to identify such novel ethical issues. For instance, researchers may consider adopting the embedded ethics approach proposed by McLennan et al [58] to investigate the ethicality of digital twins for personalized health care services with developers and stakeholders. For researchers interested in developing ethical guidance for emerging digital twin applications, the ethics parallel research approach advocated by Jongsma and Bredenoord [59] is also worth adopting.

Conclusions

The concept of digital twins can be applied to a wide variety of personalized health care services. The diversity of digital twins for personalized health care services not only manifests in the sort of health care services they aim to provide but also in the ethical risks they might face. To capture these nuances, we conducted a process-oriented ethical analysis to examine the ethical risks that could appear during data processing and information use. The 10 operational problems and relevant ethical values have been structured with a clear, logical flow. This process-oriented ethical map allows developers of digital twins for personalized health care services and stakeholders to have a comprehensive overview of major ethical risks when refining the design of the digital twin. The ethical values section on the map also helps developers better understand the values they ought to consider when developing solutions for an operational problem they might encounter.

Acknowledgments

This research was funded by the Convergence Healthcare Technology Flagship of Delft University of Technology and Erasmus MC University Medical Centre and the National Research Foundation of Korea grant funded by the Ministry of Science, ICT, and Future Planning (No. 2020R1C1C1014312).

Authors' Contributions

PH and KK were the co–first authors of this research. They contributed equally to the work.
MS provided critical feedback on the manuscript and supervised the study.

Conflicts of Interest

None declared.

References

1. Lutze R. Digital twin based software design in eHealth - a new development approach for health/medical software products. In: 2020 IEEE international conference on concurrent enterprising.: IEEE; 2020 Presented at: ICE/ITMC'20; June 15-17, 2020; Cardiff p. 1-9. [doi: 10.1109/ice/itmc49519.2020.9198546]
2. Institute of Medicine. Engineering a learning healthcare system: a look at the future: workshop summary. Washington, DC: National Academies Press; 2011.
3. Gawel DR, Serra-Musach J, Lilja S, Aagesen J, Arenas A, Asking B, et al. A validated single-cell-based strategy to identify diagnostic and therapeutic targets in complex diseases. Genome Med 2019;11(1):47. [doi: 10.1186/s13073-019-0657-3] [Medline: 31358043]
4. Dynamic HeartModel A.I.(Anatomical Intelligence) for additional insight. Philips. 2018. URL: https://www.usa.philips.com/healthcare/resources/feature-detail/ultrasound-heartmodel [accessed 2021-08-20]
5. Copley C. Medtech firms get personal with digital twins. Reuters. 2018. URL: https://www.reuters.com/article/us-healthcare-medical-technology-ai-insi-idUSKCN1LG0S0 [accessed 2021-08-20]
6. Saracco R. Digital twins: bridging physical space and cyberspace. Computer 2019;52(12):58-64. [doi: 10.1109/mc.2019.2942803]
7. Barricelli BR, Casiraghi E, Fogli D. A survey on digital twin: definitions, characteristics, applications, and design implications. IEEE Access 2019;7:167653-167671. [doi: 10.1109/access.2019.2953499]
8. Abbott D. Applied predictive analytics: principles and techniques for the professional data analyst. Hoboken: Wiley; 2014.
9. Vikhar PA. Evolutionary algorithms: a critical review and its future prospects. In: 2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication.: IEEE; 2016 Presented at: ICGTSPICC'16; December 22-24, 2016; Jalgaon p. 261-265. [doi: 10.1109/ICGTSPICC.2016.7955308]
10. van den Bemd GJ. My digital twin. Convergence Flagship. 2020. URL: https://convergencealliance.nl/my-digital-twin/ [accessed 2021-08-20]
11. Medicinsk digital tvilling (MeDigiT). Linköping University. URL: https://liu.se/forskning/medigit [accessed 2021-08-20]
12. Swedish Digital Twin Consortium. Swedish Digital Twin Consortium. 2021. URL: https://www.sdtc.se/ [accessed 2021-08-20]
13. Digital Twins for Better Health. DigiTwins. 2021. URL: https://www.digitwins.org/ [accessed 2021-01-01]

14. Ahmadi-Assalemi G, Al-Khateeb H, Maple C, Epiphaniou G, Alhaboby ZA, Alkaabi S, et al. Digital twins for precision healthcare. In: Jahankhani H, Kendzierskyj S, Chelvachandran N, Ibarra J, editors. Cyber defence in the age of AI, smart societies and augmented humanity. Cham: Springer; 2020:133-158.
15. Prainsack B. Personalized medicine: empowered patients in the 21st century?. New York: New York University Press; 2017.
16. Bruynseels K, Santoni de Sio F, van den Hoven J. Digital twins in health care: ethical implications of an emerging engineering paradigm. Front Genet 2018;9:31. [doi: 10.3389/fgene.2018.00031] [Medline: 29487613]
17. Beauchamp T, Childress J. Principles of biomedical ethics. 7th edition. New York: Oxford University Press; 2012.
18. Delen D, Demirkan H. Data, information and analytics as services. Decis Support Syst 2013;55(1):359-363. [doi: 10.1016/j.dss.2012.05.044]
19. Bagaria N, Laamarti F, Badawi HF, Albraikan A, Martinez Velazquez RA, El Saddik A. Health 4.0: digital twins for health and well-being. In: El Saddik A, Hossain MS, Kantarci B, editors. Connected health in smart cities. Cham: Springer; 2019:143-152.
20. Guidelines concerning the processing of health data in the workplace by community institutions and bodies. European Data Protection Supervisor. 2009. URL: https://edps.europa.eu/data-protection/our-work/publications/guidelines/health-data-work_en [accessed 2021-08-20]
21. Krumm J. Inference attacks on location tracks. In: Pervasive computing: 5th international conference, PERVASIVE 2007, Toronto, Canada, May 13-16, 2007. proceedings. Berlin: Springer; 2007 Presented at: Pervasive 2007; May 13-16, 2007; Toronto p. 127-143. [doi: 10.1007/978-3-540-72037-9_8]
22. de Miguel Beriain I, Morla González M. 'Digital pills' for mental diseases: an ethical and social analysis of the issues behind the concept. J Law Biosci 2020;7(1):lsaa040. [doi: 10.1093/jlb/lsaa040] [Medline: 34221420]
23. Belluck P. First digital pill approved to worries about biomedical 'Big Brother'. The New York Times. 2017. URL: https://www.nytimes.com/2017/11/13/health/digital-pill-fda.html [accessed 2021-08-20]
24. Carter A, Savic M, Forlini C. Surveillance medicine in the DigitalEra: lessons from addiction treatment. Am J Bioeth 2018;18(9):58-60. [doi: 10.1080/15265161.2018.1499832] [Medline: 30265603]
25. Swartz AK. Smart pills for psychosis: the tricky ethical challenges of digital medicine for serious mental illness. Am J Bioeth 2018;18(9):65-67. [doi: 10.1080/15265161.2018.1498948] [Medline: 30265606]
26. Armstrong D. The rise of surveillance medicine. Sociol Health Illn 1995;17(3):393-404. [doi: 10.1111/1467-9566.ep10933329]
27. Falter M, Budts W, Goetschalckx K, Cornelissen V, Buys R. Accuracy of apple watch measurements for heart rate and energy expenditure in patients with cardiovascular disease: cross-sectional study. JMIR Mhealth Uhealth 2019;7(3):e11889. [doi: 10.2196/11889] [Medline: 30888332]
28. Marcus GM. The Apple watch can detect atrial fibrillation: so what now? Nat Rev Cardiol 2020;17(3):135-136. [doi: 10.1038/s41569-019-0330-y] [Medline: 31873198]
29. Mettler T, Wulf J. Health promotion with physiolytics: what is driving people to subscribe in a data-driven health plan. PLoS One 2020;15(4):e0231705. [doi: 10.1371/journal.pone.0231705] [Medline: 32294126]
30. Sandborn PA. Editorial software obsolescence—complicating the part and technology obsolescence management problem. IEEE Trans Compon Packaging Technol 2007;30(4):886-888. [doi: 10.1109/tcapt.2007.910918]
31. Hummel P, Braun M, Dabrock P. Own data? Ethical reflections on data ownership. Philos Technol 2021;34(3):545-572. [doi: 10.1007/s13347-020-00404-9]
32. Glenn T, Monteith S. Privacy in the digital world: medical and health data outside of HIPAA protections. Curr Psychiatry Rep 2014;16(11):494. [doi: 10.1007/s11920-014-0494-4] [Medline: 25218603]
33. Lupton D. The commodification of patient opinion: the digital patient experience economy in the age of big data. Sociol Health Illn 2014;36(6):856-869. [doi: 10.1111/1467-9566.12109] [Medline: 24443847]
34. Huckvale K, Torous J, Larsen ME. Assessment of the data sharing and privacy practices of smartphone apps for depression and smoking cessation. JAMA Netw Open 2019;2(4):e192542. [doi: 10.1001/jamanetworkopen.2019.2542] [Medline: 31002321]
35. Martin G, Martin P, Hankin C, Darzi A, Kinross J. Cybersecurity and healthcare: how safe are we? BMJ 2017;358:j3179. [doi: 10.1136/bmj.j3179] [Medline: 28684400]
36. Health care and cyber security: increasing threats require increased capabilities. KPMG. 2015. URL: https://assets.kpmg.com/content/dam/kpmg/pdf/2015/09/cyber-health-care-survey-kpmg-2015.pdf [accessed 2021-08-20]
37. Hacking healthcare in 2016: lessons the healthcare industry can learn from the OPM breach 2016. Institute for Critical Infrastructure Technology. 2016. URL: http://icitech.org/wp-content/uploads/2016/01/ICIT-Brief-Hacking-Healthcare-IT-in-2016.pdf [accessed 2021-08-20]
38. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science 2019;366(6464):447-453. [doi: 10.1126/science.aax2342] [Medline: 31649194]
39. Morley J, Machado CC, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: a mapping review. Soc Sci Med 2020;260:113172. [doi: 10.1016/j.socscimed.2020.113172] [Medline: 32702587]

40. Liu C, Liu X, Wu F, Xie M, Feng Y, Hu C. Using artificial intelligence (Watson for oncology) for treatment recommendations amongst Chinese patients with lung cancer: feasibility study. J Med Internet Res 2018;20(9):e11087. [doi: 10.2196/11087] [Medline: 30257820]
41. Exploring the ChestXray14 dataset: problems. Luke Oakden-Rayner. 2017. URL: https://lukeoakdenrayner.wordpress.com/2017/12/18/the-chestxray14-dataset-problems/ [accessed 2021-08-20]
42. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S, AIX-COVNET, et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. Nat Mach Intell 2021;3:199-217. [doi: 10.1038/s42256-021-00307-0]
43. Lupton D. Digitized health promotion: risk and personal responsibility for health in the web 2.0 era. In: Davis JE, González AM, editors. To fix or to heal: patient care, public health, and the limits of biomedicine. Sydney: Sydney Health & Society Group; 2013:152-176.
44. Stol YH, Schermer MH, Asscher EC. Omnipresent health checks may result in over-responsibilization. Public Health Ethics 2017;10(1):35-48. [doi: 10.1093/phe/phw034]
45. Owens J, Cribb A. 'My Fitbit thinks I can do better!' Do health promoting wearable technologies support personal autonomy? Philos Technol 2019;32:23-38. [doi: 10.1007/s13347-017-0266-2]
46. Marantz PR. Blaming the victim: the negative consequence of preventive medicine. Am J Public Health 1990;80(10):1186-1187. [doi: 10.2105/ajph.80.10.1186] [Medline: 2400027]
47. Rich E, Lewis S, Lupton D, Miah A, Piwek L. Digital health generation? Young people's use of 'healthy lifestyle' technologies. Bath: University of Bath; 2020.
48. Ruckenstein M, Schüll ND. The datafication of health. Annu Rev Anthropol 2017;46:261-278. [doi: 10.1146/annurev-anthro-102116-041244]
49. Walker MJ, Rogers W. Defining disease in the context of overdiagnosis. Med Health Care Philos 2017;20(2):269-280. [doi: 10.1007/s11019-016-9748-8] [Medline: 27848107]
50. Mandl KD, Manrai AK. Potential excessive testing at scale: biomarkers, genomics, and machine learning. JAMA 2019;321(8):739-740. [doi: 10.1001/jama.2019.0286] [Medline: 30735228]
51. Bunnik EM, Janssens AC, Schermer MH. Personal utility in genomic testing: is there such a thing? J Med Ethics 2015;41(4):322-326. [doi: 10.1136/medethics-2013-101887] [Medline: 24872596]
52. U.S. Preventive Services Task Force. Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: recommendation statement. Ann Intern Med 2005;143(5):355-361. [doi: 10.7326/0003-4819-143-5-200509060-00011] [Medline: 16144894]
53. Katz SJ, Morrow M. Addressing overtreatment in breast cancer: the doctors' dilemma. Cancer 2013;119(20):3584-3588. [doi: 10.1002/cncr.28260] [Medline: 23913512]
54. Kingma E. Paracetamol, poison, and polio: why Boorse's account of function fails to distinguish health and disease. Br J Philos Sci 2010;61(2):241-264. [doi: 10.1093/bjps/axp034]
55. Nickel B, Moynihan R, Barratt A, Brito JP, McCaffery K. Renaming low risk conditions labelled as cancer. BMJ 2018;362:k3322. [doi: 10.1136/bmj.k3322] [Medline: 30100549]
56. Chadwick R, Levitt M, Schickle D. The right to know and the right not to know: genetic privacy and responsibility. 2nd edition. Cambridge: Cambridge University Press; 2014.
57. O'Neill O. Informed consent and genetic information. Stud Hist Philos Biol Biomed Sci 2001;32(4):689-704. [doi: 10.1016/s1369-8486(01)00026-7]
58. McLennan S, Fiske A, Celi LA, Müller R, Harder J, Ritt K, et al. An embedded ethics approach for AI development. Nat Mach Intell 2020;2:488-490. [doi: 10.1038/s42256-020-0214-1]
59. Jongsma KR, Bredenoord AL. Ethics parallel research: an approach for (early) ethical guidance of biomedical innovation. BMC Med Ethics 2020;21(1):81. [doi: 10.1186/s12910-020-00524-z] [Medline: 32867753]

Abbreviations

ICT: information and communication technology
MRI: magnetic resonance imaging

Edited by K El Emam; submitted 23.08.21; peer-reviewed by J Sharp, A Power; comments to author 19.10.21; revised version received 27.10.21; accepted 16.11.21; published 31.01.22.

Please cite as:
Huang PH, Kim KH, Schermer M
Ethical Issues of Digital Twins for Personalized Health Care Service: Preliminary Mapping Study
J Med Internet Res 2022;24(1):e33081
URL: https://www.jmir.org/2022/1/e33081
doi:10.2196/33081
PMID:35099399

Original Paper

Characterizing Awareness of Pre-Exposure Prophylaxis for HIV Prevention in Manila and Cebu, Philippines: Web-Based Survey of Filipino Cisgender Men Who Have Sex With Men

Arjee Restar[1,2,3,4,5,6*], MPH, PhD; Anthony Surace[1*], MA; Alexander Adia[1,2], MPH; William Goedel[1], PhD; Adedotun Ogunbajo[1], MHS, MPH, PhD; Harry Jin[1,4], MPH, PhD; Alberto Edeza[1], PhD; Laufred Hernandez[7], MMPM, MA; Susan Cu-Uvin[1,2,8,9], MD; Don Operario[1,2], PhD

[1]Department of Behavioral and Social Sciences, Brown University School of Public Health, Providence, RI, United States

[2]Brown University Global Health Initiative, The Philippine Health Initiative for Research, Service, and Training, Providence, RI, United States

[3]amFAR, The Foundation of AIDS Research, Washington, DC, United States

[4]Department of Epidemiology, Bloomberg School of Public Health, Johns Hopkins University, Baltimore, MD, United States

[5]Department of Epidemiology, School of Public Health, University of Washington, Seattle, WA, United States

[6]Department of Epidemiology, Brown University School of Public Health, Providence, RI, United States

[7]Department of Behavioral Sciences, University of Philippines in Manila, Manila, Philippines

[8]Providence-Boston Center for AIDS Research, Providence, RI, United States

[9]Department of Medicine, Miriam Hospital, Providence, RI, United States

[*]these authors contributed equally

Corresponding Author:
Arjee Restar, MPH, PhD
Department of Epidemiology
School of Public Health
University of Washington
3980 15th Ave NE
Seattle, WA, 98195
United States
Phone: 1 7046409953
Email: arjee_restar@alumni.brown.edu

Abstract

Background: The Philippines is experiencing an HIV crisis and is considering implementing pre-exposure prophylaxis (PrEP) as a national public health strategy for HIV prevention for cisgender men who have sex with men (cis-MSM). However, critical information on the awareness of PrEP among cis-MSM is needed to roll out this public health initiative.

Objective: This study aims to assess PrEP awareness and related correlates (ie, sociodemographic variables, social factors, and health care access and use) among Filipino cis-MSM.

Methods: We conducted a web-based survey with Filipino cis-MSM (n=179) residing in the cities of Manila and Cebu, Philippines. Multivariable analysis procedures were performed to examine the factors associated with PrEP awareness.

Results: Our sample demonstrated high awareness (134/179, 74.9%) and interest (159/179, 88.8%) in taking PrEP. The adjusted model showed that greater odds of PrEP awareness were associated with having a college education or higher versus a high school education or lower (adjusted odds ratio [aOR] 7.30, 95% CI 1.01-52.47), earning between PHP 10,000 (US $198.6) and PHP 20,000 (US $397.2) versus <PHP 10,000 (US $198.60; aOR 9.32, 95% CI 1.41-6.22), having had a prior HIV test (aOR 6.06, 95% CI 1.20-13.55), having high HIV knowledge (aOR 3.50, 95% CI 1.11-10.98), and having friends who discussed PrEP (aOR 11.17, 95% CI 2.73-14.5).

Conclusions: Our findings demonstrate that Filipino cis-MSM are aware of and interested in taking PrEP, but there is currently an unmet need for such biomedical HIV prevention technologies among this population. Incorporating PrEP education into routine HIV screening and leveraging cis-MSM social networks may be useful in optimizing potential PrEP implementation in the Philippines.

(J Med Internet Res 2022;24(1):e24126) doi:10.2196/24126

KEYWORDS

HIV prevention; PrEP; men who have sex with men; the Philippines

Introduction

Background

Globally, cisgender men (ie, individuals who identify their gender as men and are assigned the male sex at birth) who have sex with men (cis-MSM) remain highly vulnerable to HIV infection [1,2]. A meta-analysis found that the HIV prevalence among cis-MSM in the South and Southeast Asia region was 15% [2]. The Philippines, situated in Southeast Asia, is undergoing a major public health crisis as a result of the rapid increase in new HIV cases documented in the past decade. The number of people diagnosed with HIV in the Philippines has increased more than 4-fold, from 16,000 cases in 2010 to 68,000 in 2017 [3] and is concentrated among young, cis-MSM [4]. Of the 6011 new cases of HIV in 2014, 85% were attributed to male-to-male sexual contact with a median age at diagnosis of 28 years [4,5]. The major drivers of HIV infection among Filipino cis-MSM include receptive anal intercourse and sex under the influence of alcohol and other substances [5,6]. Use of HIV prevention services among this population remains low; in a survey of 531,000 cis-MSM in the Philippines, only half reported using condoms during their last sexual encounter [3], and HIV testing rates were lower among cis-MSM compared with those in China and Vietnam [7].

Pre-exposure prophylaxis (PrEP) is an evidence-based, effective biomedical HIV prevention strategy that significantly reduces HIV transmission risk [8]. Currently, PrEP is available as a daily oral pill (eg, Truvada, a combination of emtricitabine and tenofovir), which can reduce the risk of HIV infection by more than 90% when taken as prescribed [9,10]. To date, there is a single ongoing PrEP demonstration project among cis-MSM in the Philippines [11], and the Philippines Department of Health is considering approval of PrEP for public use. Examining levels of PrEP awareness and interest among cis-MSM in the Philippines will inform future rollout and scaling of PrEP programs, especially for vulnerable groups such as cis-MSM.

Awareness of and interest in taking PrEP are key determinants of successful initiation into the PrEP continuum of care [12,13]. PrEP awareness has been researched among American cis-MSM [14,15], but analogous research in other parts of the world is scant. One of the few published reviews examining awareness and willingness to use PrEP among cis-MSM in low- and middle-income countries found low awareness (16.9%-44.3%) but high willingness to use PrEP for HIV prevention (53.3%-74.8%) [16]. Furthermore, a literature review of HIV research in the Philippines showed that there are only a few empirical studies on HIV prevention among cis-MSM, and none of these studies have explored PrEP [17]. Given the ongoing considerations to support implementation of PrEP in national guidelines in the Philippines, it is vital to assess the current level of PrEP awareness and interest among Filipino cis-MSM. Such data are critical for informing future PrEP implementation programs in the Philippines.

Objective

To date, no known quantitative study has explored awareness of PrEP among Filipino cis-MSM. There is a need for such research given the high vulnerability of Filipino cis-MSM to HIV infection and the heavy burden of HIV infection among this group. To this end, we carried out a web-based survey assessment to examine the relationship between PrEP awareness among Filipino cis-MSM and sociodemographic variables, social factors (eg, marginalization, cohesion, and participation), indicators of health care experiences and access to HIV services, and PrEP indicators.

Methods

Participants and Procedures

The data used in this study were obtained from Project #ParaSaAtin, a web-based study that surveyed key populations (eg, cis-MSM) impacted by the HIV epidemic in the Philippines. Informed by epidemiological data, we focused on Manila or National Capital Region (NCR) and Cebu. Specifically, as identified in the Philippines HIV national surveillance report, these are the top geographical areas where HIV infection is concentrated in the Philippines [4]. The purpose of the survey was to examine the structural, social, and behavioral factors that impact the use of HIV/AIDS prevention services, such as PrEP, condoms, and HIV testing. This analysis focused specifically on PrEP awareness and willingness to take PrEP as outcomes of interest.

Between June 2018 and May 2019, we recruited participants by posting our study survey link on flyers and advertisements through web-based social media sites (eg, Facebook groups and Twitter) of community-based organizations that provided services to communities impacted by the HIV/AIDS epidemic in the Philippines. Participants who clicked on the survey completed a web-based written consent and screening process. Eligibility criteria for participants in this study included self-reporting that they (1) were ≥18 years old, (2) identified as cis-MSM, (3) had condomless anal sex within the past year, (4) currently lived in Manila, NCR or Cebu, (5) demonstrated English comprehension (via a brief cognitive screening form), and (6) provided a web-based written consent to participate in the study. The brief cognitive screening form tested participants' capacity to understand English by asking participants a series of true or false questions based on the consent form. English is common in the Philippines (as it is one of the national languages) [18] but given that the survey was written in English, we wanted to ensure that the enrolled participants demonstrated sufficient language comprehension. To further ensure survey comprehension, we used the Flesch-Kincaid Reading Level Test [19] to design the survey such that it could be read by someone at the sixth grade level. The average range of time for participants to complete the survey was between 20 and 25 minutes. All participants received a PHP 300 (approximately

US $5.85) electronic prepaid mobile load card for their time and completion of the survey.

We also used best practices for conducting web-based survey questionnaires. This included having a series of standard web-based survey protective measures, such as (1) *captcha box*, which provides a web-based challenge-response test that eliminates robots from taking the survey [20,21] and (2) unique IP address programming, which prevents the same participants from taking the survey multiple times by only allowing those with a unique IP address to take the survey [21,22]. Study procedures were deemed appropriate by all 3 local HIV/AIDS community-based organizations that were our local partners for this project and collaborated in the development and implementation of this survey. This study received ethical approval from the Brown University Human Research Protection Program Institutional Review Committee in Providence, Rhode Island.

Measures

We analyzed the correlates of PrEP awareness and interest using a series of questions on sociodemographic factors, social factors (eg, marginalization, cohesion, and participation), and indicators of health care experiences and access to HIV services among the participants.

Sociodemographics

We adapted sociodemographic items from the Philippines National Demographic and Health Survey [23]. With the exception of sexual identity, these items included age, current living location, education attainment, past year income, and religious affiliation.

Social Marginalization, Cohesion, and Participation Variables

We asked the participants about their social experiences, including marginalization, cohesion (ie, perception of mutual aid, trust, and support) [24], and participation in social events and activities.

For social marginalization, we asked if they had ever been homeless (*yes or no*), were currently unemployed (*yes or no*), and had engaged in sex work within the past 4 months (*yes or no*).

To measure social cohesion, we adapted a 9-item social cohesion scale [24] to specifically refer to cis-MSM communities (Cronbach α=.92). Examples of these items are *You can count on your cis-MSM friends if you need to talk about your problems* and *You can count on your cis-MSM friends to accompany you to the doctor or hospital*. Each item had 5-level Likert response options, from *strongly disagree* (1) to *strongly agree* (5) and was summed (median 22; range 9-45). Similar to previous researchers' work on this scale [24-26], scores were dichotomized into low versus high levels of social cohesion at the median.

We asked about the engagement of the participants in general community events and activities (ie, general social participation), such as activities related to church, community, or cultural events as well as lesbian, gay, bisexual, and transgender (LGBT)-specific activities (ie, LGBT-specific social participation), such as joining pageants, HIV advocacy groups, and local LGBT organizations. To assess gender- and LGBT-specific social participation, we adapted the social participation scale (Cronbach α=.76 and .75, respectively) [25]. Keeping in-line with previous researchers [25,26], responses were either *yes* or *no* and were summed (median 1, range 0-4 for general social participation and median 3, range 0-5 for LGBT-specific social participation) and then dichotomized into low versus high levels of participation at the median.

Health Care and HIV Service Indicators

We assessed participants' health care experience by asking whether they had current health insurance (*yes or no*) and had ever been discriminated against because of sexual identity (*yes or no*). We used the health care accessibility scale by Haggerty et al [27] to measure our participants' health care accessibility (Cronbach α=.89). Items included questions about how convenient their provider's office is with regard to office hours for appointments, duration of next open appointments, waiting time before appointments, traveling to and from the office, and ability to reach the provider when needed. Responses were scored (median 19, range 6-30) and dichotomized into poor or fair versus good or excellent health care accessibility at the median.

We assessed the experiences of access to HIV services, by asking the participants if they had ever had an HIV test. We also asked if they avoided seeking HIV services because of the cost of services, distance of travel to and from the health care facility, sexual identity, and lack of facility-level policy on LGBT antidiscrimination. To assess participant's HIV knowledge, we used the International AIDS Questionnaire (Cronbach α=.80; median 40, range 23-74), and dichotomized the scores into high versus low HIV knowledge at the median [28].

PrEP Indicators

We adapted 3 items on PrEP awareness, interest, and discussion among friends from the study by Restar et al [29]. Specifically, we asked the participants whether they had heard of PrEP (*yes or no*). Regardless of their previous PrEP awareness, all participants then received a brief educational statement about PrEP [30]:

> *One way to fight HIV is called PrEP, which stands for pre - exposure prophylaxis. PrEP works by giving HIV - negative people HIV drugs to keep them from getting HIV. The following questions are about your thoughts and opinions of PrEP.*

We then asked if they were interested in taking PrEP (*not interested at all* vs *somewhat or very interested*). We also asked the participants to rate how true the following statement is—*My friends talk about PrEP*—with response options as *very or somewhat true* versus *not true at all*. Participants who were not interested in PrEP were subsequently asked to enter their primary reason for being uninterested in PrEP.

Analysis

Descriptive analyses were performed to determine proportions and chi-square tests and bivariable and multivariable logistic regression modeling were performed to determine associations among participants. Before running the regression procedures, we performed a sensitivity analysis to determine internal consistency for all of our scale variables (eg, health care accessibility, social cohesion and participation, and HIV knowledge). Multicollinearity tests using the standard procedure for assessing the variance inflation factor were performed to ensure that our independent variables were not collinear. Our results determined that collinearity was not present in our data (all variance inflation factors <4.00) [31]. Logistic regression procedures were performed to determine associations between PrEP awareness and sociodemographics, experiences of social marginalization, social cohesion and participation, health care, access to HIV services, and other PrEP indicators. We then descriptively examined the reasons for not having interest in PrEP. All analyses were performed using Stata standard edition version 15.1 (StataCorp LLC) [32]. Statistical significance was set at 2-sided $P<.05$.

Results

Sample Characteristics

A total of 179 cis-MSM completed the survey. Study sample characteristics are presented in (Table 1).

Table 1. Study sample characteristics of Filipino cismen who have sex with men (n=179).

Characteristics	Total, n (%)	PrEP[a]-aware (n=134), n (%)	*P* value[b]
Demographics			
Age			.08
18-24	40 (23)	32 (24.2)	
25-29	59 (33.9)	45 (34.1)	
30-34	36 (20.7)	31 (23.5)	
>35	39 (22.4)	24 (18.2)	
Current living location			.52
Metro Manila or National Capital Region	145 (81)	110 (82.1)	
Central Visayas	34 (19)	24 (17.9)	
Highest educational attainment			.10
High school or lower	23 (12.9)	13 (9.7)	
Some college	34 (19.1)	26 (19.6)	
College or higher	121 (68)	94 (70.7)	
Past year income			.27
<PHP 10,000 (US $198.60)	25 (19.6)	21 (15.7)	
PHP 10,000 (US $198.60)-PHP 20,000 (US $397.20)	38 (21.2)	30 (22.4)	
PHP 20,000 (US $397.20)-PHP 30,000 (US $ 596.00)	30 (16.8)	23 (17.2)	
>PHP 30,000 (US $ 596.00)	60 (33.5)	47 (35)	
No income in the past year	16 (8.9)	13 (9.7)	
Religious affiliation			.14
Catholic	144 (80.4)	105 (78.4)	
Non-Catholic (eg, Protestant, Christian)	24 (13.4)	18 (13.4)	
Nonreligious	11 (6.2)	11 (8.2)	
Sexual orientation			.048[c,d]
Gay	95 (53)	76 (56.7)	
Bisexual	71 (39.7)	52 (38.8)	
Straight	11 (6.2)	5 (3.7)	
Not listed	2 (1.1)	1 (0.8)	
Social marginalization, cohesion, and participation			
Ever homeless			.49
No	158 (88.3)	117 (87.3)	
Yes	21 (11.7)	17 (12.7)	
Recent (<4 months) sex work engagement			.79
No	153 (85.5)	114 (85.1)	
Yes	26 (14.5)	20 (14.9)	
Currently unemployed			.18
No	46 (25.7)	31 (23.2)	
Yes	133 (74.3)	103 (76.8)	
Social cohesion			.24
Low	82 (45.8)	58 (43.3)	
High	97 (54.2)	76 (56.7)	

Characteristics	Total, n (%)	PrEP[a]-aware (n=134), n (%)	*P* value[b]
General social participation			.09
Low	92 (51.4)	64 (47.8)	
High	87 (48.6)	70 (52.2)	
LGBT[e]-specific social participation			.002
Low	109 (60.9)	73 (54.5)	
High	70 (39.1)	61 (45.5)	
HIV and other health care indicators			
Current health insurance			.32
No	88 (49.2)	63 (47)	
Yes	91 (50.8)	71 (53)	
Health care discrimination because of sexual identity			.03[c,d]
No	153 (85.5)	110 (82.1)	
Yes	26 (14.5)	24 (17.9)	
Health care accessibility			.57
Poor or fair	93 (52)	68 (50.8)	
Good or excellent	86 (48)	66 (49.2)	
Ever had an HIV test			<.001
No	37 (20.7)	17 (12.7)	
Yes	142 (79.3)	117 (87.3)	
Avoided HIV services because of the cost of services			.006
No	107 (59.8)	88 (65.7)	
Yes	72 (40.2)	46 (34.3)	
Avoided HIV services because of distance of travel to and from the health care facility			.70
No	107 (59.8)	79 (59)	
Yes	72 (40.2)	55 (41)	
Avoided HIV services because of sexual identity			.04
No	129 (72.1)	102 (76.1)	
Yes	50 (27.9)	32 (23.9)	
Avoided HIV services because of lack of LGBT antidiscrimination policy			.03
No	143 (79.9)	112 (83.6)	
Yes	36 (20.1)	22 (16.4)	
HIV knowledge			.004
Low	78 (43.6)	50 (37.3)	
High	101 (56.4)	84 (62.7)	
PrEP-related indicators			
PrEP discussion among friends			<.001
No	103 (57.5)	63 (47)	
Yes	76 (42.5)	71 (53)	
PrEP interest			.57
Not at all	20 (11.2)	16 (11.9)	

Characteristics	Total, n (%)	PrEP[a]-aware (n=134), n (%)	*P* value[b]
Very or somewhat	159 (88.8)	118 (88.1)	

[a]PrEP: pre-exposure prophylaxis.

[b]Chi-square test.

[c]Fisher exact test. Comparison groups for both tests were pre-exposure prophylaxis–aware versus pre-exposure prophylaxis–unaware.

[d]*P*<.05.

[e]LGBT: lesbian, gay, bisexual, and transgender.

Overall, most of the participants currently lived in Manila or NCR (145/179, 81%), had at least college education (121/179, 67.6%), self-identified as Catholic (144/179, 80.4%), and self-identified as gay (95/179, 53%). Approximately one-third of the participants earned PHP 30,000 or more a year (60/179, 33.5%) and were between ages 25 and 29 years (59/179, 33.9%). Most had never experienced homelessness (158/179, 88.3%) and had not recently been involved in sex work (153/179, 85.5%), but most were currently unemployed (133/179, 74.3%). In addition, approximately half of the participants displayed high social cohesion (97/179, 54.2%), showed high general social participation (87/179, 48.6%), exhibited low LGBT-specific social participation (109/179, 60.9%), and reported poor or fair health care accessibility (93/179, 52%).

Half of the participants reported having no current health insurance (88/179, 49.2%), and 14.5% (26/179) reported having experienced health care discrimination because of their sexual identity. For HIV indicators, most of the participants had taken an HIV test (142/179, 79.3%), and less than half had avoided HIV services because of the cost of services (72/179, 40.2%), distance of travel to and from the health care facility (72/179, 40.2%), sexual identity (50/179, 27.9%), and lack of LGBT antidiscrimination policy (36/179, 20.1%). Over half of the participants demonstrated having high HIV knowledge (101/179, 56.4%).

Most of the participants were aware of PrEP (134/179, 74.9%) and were interested in taking PrEP (159/179, 88.8%), and approximately half discussed PrEP among friends (76/179, 42.5%). In total, 11.7% (20/179) of Filipino cis-MSM indicated that they were not interested in taking PrEP and provided their primary reasons for not being interested. Of this subsample, the most common reasons for not being interested in PrEP were the *need to know more information before taking PrEP* (5/20, 25%) and *don't like taking pills* (5/20, 25%).

Bivariable Logistic Regression Findings

Findings from bivariable logistic regression on the correlates of PrEP awareness are presented in Multimedia Appendix 1.

Factors that were associated with greater odds of PrEP awareness included (*P*<.05 in all instances) the following: having attained college education or higher versus high school education or lower (odds ratio [OR] 2.67, 95% CI 1.05-6.77), having high LGBT-specific social participation (OR 3.34, 95% CI 1.49-7.48), having had an HIV test (OR 5.50, 95% CI 2.53-11.98), having high HIV knowledge (OR 2.76, 95% CI 1.37-5.55), and having discussed PrEP among friends (OR 9.01, 95% CI 3.35-24.25).

Factors that were associated with lower odds of PrEP awareness included (*P*<.05 in all instances) the following: being straight-identified (OR 0.20, 95% CI 0.05-0.75) and having experienced health care discrimination because of sexual identity (OR 0.21, 95% CI 0.04-0.94). In addition, avoiding HIV services because of cost (OR 0.38, 95% CI 0.19-0.76), sexual identity (OR 0.47, 95% CI 0.22-0.96), and lack of LGBT antidiscrimination policy (OR 0.43, 95% CI 0.19-0.94) were all associated with lower odds of PrEP awareness.

Multivariable Logistic Regression Findings

The findings from the multivariable logistic regression model are presented in Multimedia Appendix 1. Significantly greater odds of PrEP awareness were found among Filipino cis-MSM who had attained college education or higher versus high school education or lower (adjusted OR [aOR] 7.30, 95% CI 1.01-52.47) and earned between PHP 10,000 and PHP 20,000 versus <PHP 10,000 (aOR 9.32, 95% CI 1.41-6.22). Having ever had an HIV test was associated with approximately 6 times the odds of PrEP awareness (aOR 6.06, 95% CI 1.20-13.55) compared with not having had an HIV test. Those who displayed high HIV knowledge had more than 3 times the odds of being aware of PrEP compared with those with low HIV knowledge (aOR 3.50, 95% CI 1.11-10.98). Filipino cis-MSM who discussed PrEP among friends had greater odds of PrEP awareness (aOR 11.17, 95% CI 2.73-14.50) compared with those who did not.

Significantly lower odds of PrEP awareness were associated with age and location. Specifically, participants aged 25-29 years (aOR 0.15, 95% CI 0.02-0.90) and ≥35 years (aOR 0.11, 95% CI 0.01-0.84) had lower PrEP awareness compared with 18- to 24-year-old participants. Participants living in Cebu had lower odds of PrEP awareness (aOR 0.83, 95% CI 10.70-0.98) compared with those living in Metro Manila or NCR region.

Discussion

Overview

This paper offers insights on the potential of PrEP as an HIV prevention strategy among cis-MSM in the Philippines, a country with one of the fastest growing HIV crises in the world. Currently, the Philippines does not have an established PrEP program and is still in the nascent stages of rollout with only 1 current PrEP demonstration project. This follows broader trends in Asia, where few countries have implemented PrEP programs [33]. Insights in this paper point to a growing need and demand for PrEP in the Philippines for populations highly impacted by HIV, such as cis-MSM populations. The potential for the rollout of PrEP-based interventions and programming is contingent on

public health efforts to increase levels of PrEP awareness of and interest in taking PrEP among cis-MSM in the Philippines.

Overall, PrEP awareness and interest in taking PrEP were high in our sample, as most participants indicated that they were both aware of and interested in PrEP. There have been few efforts to document PrEP awareness and interest across the Asia-Pacific region; what little research exists suggests that knowledge about PrEP is low in the region [33]. Our findings, however, contradict such research, as we found that most participants were aware of PrEP and had a high degree of PrEP knowledge. These results may have been because of our use of community-based organizations, which had HIV-specific programming and content, for recruitment. Moreover, previous research has found associations between high HIV knowledge and having access to and better engagement with HIV services [34,35]. In our sample, HIV knowledge was associated with greater odds of PrEP awareness. As such, it is plausible that Filipino cis-MSM may gain HIV knowledge and awareness of PrEP through their connections to HIV prevention services. However, further research is needed to elucidate this connection. Ultimately, this points to a potential role for community-based organizations in generating demand for and education on PrEP, a role that organizations in other settings have been willing and able to take.[36]. Indeed, Republica Act 11166, the new HIV policy recently passed in the Philippines, explicitly calls out the need for policy makers, practitioners, and scientists to work alongside such organizations in building a stronger HIV response [37]. Given these findings, a key policy step in building PrEP rollout in the Philippines could be the integration of community-based organizations as partners in the strategic planning and implementation processes from the very start of the national PrEP program.

The results from our multivariable model also imply the means by which PrEP implementation may be optimized. First, the association between having ever had an HIV test and awareness of PrEP is positive, as it demonstrates the potential linkage of PrEP education with routine health care, such as HIV tests. Prior research has argued that this routinization of PrEP discussions in the provision of other health services both destigmatizes PrEP and facilitates broader PrEP awareness [38]. As the Philippines continues to design the implementation of its own PrEP program, the integration of the burgeoning PrEP program alongside the continued expansion of HIV testing and service provision could yield the strongest uptake.

Second, the finding that discussion of PrEP with friends was linked to greater awareness of PrEP points to how social networks may be leveraged to improve PrEP uptake. Prior research has noted that social networks may play a role in promoting PrEP awareness [39]. However, research has also shown that MSM social networks may be a source of PrEP stigma—namely that PrEP is associated with sexual promiscuity and sexually transmitted infections [40]. Such stigma may serve as a barrier to PrEP education and use [41]. Although we did not specifically assess the nature of the discussions about PrEP in this study, future research should aim to do so to gauge the feasibility of using extended networks of cis-MSM to spread PrEP awareness and increase demand and use. Questions about the directionality of how knowledge is passed on (eg, whether having awareness of PrEP may lead to more discussion among friends about PrEP or whether having friends who discuss PrEP may lead to increased awareness of PrEP) and how individuals seek information about PrEP among friends who are also informed about PrEP within their social circles remain unanswered. Qualitative research may be especially useful in exploring this, as it may allow for a deeper understanding of the complex dynamics of social interaction around PrEP. Such research may shed light on how discussing PrEP with friends impacts awareness and desire to use PrEP.

Finally, it is worth noting the other correlates of PrEP awareness levels that we found. PrEP awareness in this sample was significantly lower among those over the age of 25 years (compared with those who were younger), who had very low income and education levels (compared with those with moderate to high income or education), and who resided in Cebu (compared with Manila). These results suggest that there are distinct health disparities across regions and socioeconomic strata within the population of cis-MSM in the Philippines. In addition to sociodemographics, it is important for future research in this area to better understand the sexual behaviors and relationship contexts (eg, casual vs romantic, long-term vs short-term, and monogamous vs having more than one partner) of cis-MSM who are unaware of PrEP, as PrEP may likely be beneficial to them and their partners. Such research can better inform PrEP campaigns aimed at promoting equity across cis-MSM populations. As such, there is a need for continually monitoring the reach of PrEP education and communication campaigns to mitigate such disparities within the country. Informational campaigns must particularly acknowledge differences in language or dialect, region, and literacy levels as potential barriers to PrEP awareness.

Principal Findings

Filipino cis-MSM are aware of and interested in taking PrEP. Those who were less aware of PrEP tended to come from poorer backgrounds and have less education. In addition, those who had encountered barriers to health care (ie, costs of medical visits or discrimination from health care staff) were less likely to be aware of PrEP. Incorporating PrEP education into routine HIV screening and leveraging cis-MSM social networks may be useful in increasing PrEP awareness and uptake.

Limitations

This study has several limitations worth noting. First, as we relied on sampling through community-based organizations, our sample is unlikely to be representative of the cis-MSM populations and therefore, our findings cannot be generalized to all cis-MSM in the Philippines, particularly those who may be at higher risk for HIV. That is, although all enrolled participants reported condomless anal sex in the past year, our participants come from 2 urban areas and most of them were educated (ie, approximately 121/179, 67.6% had completed college education or above and 34/179, 19% had some college education) and therefore, may be more aware of HIV than those who do not have access to community-based organizations that offer HIV-related programs. As such, our findings may not accurately reflect the population of high-risk cis-MSM. This is underscored by the high level of HIV knowledge in our sample.

Second, we recruited participants from major metropolitan areas in the Philippines, so our findings likely do not capture awareness and education in other regions (eg, rural areas). Finally, our survey was only offered in English; although English is a national language of the Philippines, those only comfortable communicating in Tagalog (or other dialects in the Philippines) were screened out. These individuals may differ significantly from our study sample (eg, comfort communicating only in Tagalog may reflect lower access to education) in ways that may have impacted our results.

Conclusions

Our research represents the first effort to characterize PrEP awareness and willingness to take PrEP among a sample of Filipino cis-MSM. Our findings demonstrate that Filipino cis-MSM are aware of and interested in taking PrEP, but there is currently an unmet need for such biomedical HIV prevention technologies among this population. However, it must also be acknowledged that participants in our sample were likely to be engaged with HIV prevention services. More work is needed to identify the needs of less visible Filipino cis-MSM (ie, those disconnected from HIV prevention services who may be underserved by the national health care system of the Philippines). This study contributes to the scant research on this population and offers insights on the potential of PrEP as a means to reduce HIV among this population.

Acknowledgments

The authors would like to thank the health care providers who participated in this study. Their appreciation goes out to members of the study team: Patricia Rodarte, Savannah Gomes, Bianca Obiakor, and Valerie Santos. This work was supported by the Fogarty International Center of the National Institutes of Health (NIH) under grant D43TW010565, the National Institute on Drug Abuse of the NIH under grant R36DA048682, the Providence and Boston Center for AIDS Research under grant P30AI042853, the National Institute of Allergy and Infectious Diseases under grant T32AI102623, and the National Institute on Minority Health and Health Disparities of the NIH under grant 5T37MD008655 at the Brown University Global Health Initiative. AR is a recipient of the Robert Wood Johnson Foundation Health Policy Research Scholars and a Public Policy Fellow at amfAR, the Foundation for AIDS Research. The views and opinions expressed in this paper are solely those of the authors and do not necessarily represent the official views of the sponsors. The Fogarty International Center, National Institute on Drug Abuse, Providence and Boston Center for AIDS Research, National Institute of Allergy and Infectious Diseases, National Institute on Minority Health and Health Disparities, Robert Wood Johnson Foundation, and amfAR had no role in the design of the study; in the collection, analysis, and interpretation of the data; or in writing the manuscript.

Authors' Contributions

AR, AS, AA, WC, AO, HJ, and AE analyzed and interpreted the data regarding pre-exposure prophylaxis awareness. AR, AS, AA, WC, AO, HJ, and AE prepared the manuscript and were major contributors in writing the manuscript. AR, LH, SCU, and DO designed the study and were major contributors in writing, reviewing, and editing the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Bivariable and multivariable logistic regression models.
[DOCX File , 22 KB - jmir_v24i1e24126_app1.docx]

References

1. Beyrer C, Sullivan P, Sanchez J, Baral SD, Collins C, Wirtz AL, et al. The increase in global HIV epidemics in MSM. AIDS 2013 Nov 13;27(17):2665-2678. [doi: 10.1097/01.aids.0000432449.30239.fe] [Medline: 23842129]
2. Beyrer C, Baral SD, van Griensven F, Goodreau SM, Chariyalertsak S, Wirtz AL, et al. Global epidemiology of HIV infection in men who have sex with men. Lancet 2012 Jul 28;380(9839):367-377 [FREE Full text] [doi: 10.1016/S0140-6736(12)60821-6] [Medline: 22819660]
3. Nygren-Krug H. The Joint United Nations Programme on HIV/AIDS. In: Human Rights in Global Health: Rights-Based Governance for a Globalizing World. Oxford, UK: Oxford Scholarship Online; 2018.
4. Department of Health Epidemiology Bureau. 2019. URL: https://doh.gov.ph/sites/default/files/statistics/EB_HARP_September_AIDSreg2019.pdf [accessed 2021-11-17]
5. Ross AG, Ditangco RA, Belimac JG, Olveda RM, Mercado ES, Rogers GD, et al. HIV epidemic in men who have sex with men in Philippines. Lancet Infect Dis 2013 Jun;13(6):472-473. [doi: 10.1016/S1473-3099(13)70129-4] [Medline: 23718916]
6. Gangcuangco LM, Tan ML, Berba RP. Prevalence and risk factors for HIV infection among men having sex with men in Metro Manila, Philippines. Southeast Asian J Trop Med Public Health 2013 Sep;44(5):810-817. [Medline: 24437316]

7. Toleran DE, Friese B, Battle RS, Gardiner P, Tran PD, Lam J, et al. Correlates of HIV and HCV risk and testing among Chinese, Filipino, and Vietnamese men who have sex with men and other at-risk men. AIDS Educ Prev 2013 Jun;25(3):244-254. [doi: 10.1521/aeap.2013.25.3.244] [Medline: 23631718]
8. Consolidated Guidelines on HIV Prevention, Diagnosis, Treatment and Care for Key Populations. Geneva: World Health Organization; 2014.
9. Grant RM, Lama JR, Anderson PL, McMahan V, Liu AY, Vargas L, iPrEx Study Team. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. N Engl J Med 2010 Dec 30;363(27):2587-2599 [FREE Full text] [doi: 10.1056/NEJMoa1011205] [Medline: 21091279]
10. Choopanya K, Martin M, Suntharasamai P, Sangkum U, Mock PA, Leethochawalit M, Bangkok Tenofovir Study Group. Antiretroviral prophylaxis for HIV infection in injecting drug users in Bangkok, Thailand (the Bangkok Tenofovir Study): a randomised, double-blind, placebo-controlled phase 3 trial. Lancet 2013 Jun 15;381(9883):2083-2090. [doi: 10.1016/S0140-6736(13)61127-7] [Medline: 23769234]
11. Philippines. PrEPWatch. URL: https://www.prepwatch.org/country/philippines/) [accessed 2021-11-17]
12. Kelley CF, Kahle E, Siegler A, Sanchez T, Del Rio C, Sullivan PS, et al. Applying a PrEP continuum of care for men who have sex with men in Atlanta, Georgia. Clin Infect Dis 2015 Nov 15;61(10):1590-1597 [FREE Full text] [doi: 10.1093/cid/civ664] [Medline: 26270691]
13. Nunn A, Brinkley-Rubinstein L, Oldenburg CE, Mayer KH, Mimiaga M, Patel R, et al. Defining the HIV pre-exposure prophylaxis care continuum. AIDS 2017 Mar 13;31(5):731-734 [FREE Full text] [doi: 10.1097/QAD.0000000000001385] [Medline: 28060019]
14. Goedel WC, Halkitis PN, Greene RE, Duncan DT. Correlates of awareness of and willingness to use Pre-exposure Prophylaxis (PrEP) in gay, bisexual, and other men who have sex with men who use geosocial-networking smartphone applications in New York City. AIDS Behav 2016 Jul;20(7):1435-1442. [doi: 10.1007/s10461-016-1353-6] [Medline: 26966013]
15. Bauermeister J, Meanley S, Pingel E, Soler J, Harper G. PrEP awareness and perceived barriers among single young men who have sex with men. Curr HIV Res 2013 Oct;11(7):520-527 [FREE Full text] [doi: 10.2174/1570162x12666140129100411] [Medline: 24476355]
16. Yi S, Tuot S, Mwai GW, Ngin C, Chhim K, Pal K, et al. Awareness and willingness to use HIV pre-exposure prophylaxis among men who have sex with men in low- and middle-income countries: a systematic review and meta-analysis. J Int AIDS Soc 2017 Jun 26;20(1):21580 [FREE Full text] [doi: 10.7448/IAS.20.1.21580] [Medline: 28691439]
17. Restar A, Nguyen M, Nguyen K, Adia A, Nazareno J, Yoshioka E, et al. Trends and emerging directions in HIV risk and prevention research in the Philippines: a systematic review of the literature. PLoS One 2018 Dec 5;13(12):e0207663 [FREE Full text] [doi: 10.1371/journal.pone.0207663] [Medline: 30517178]
18. Gonzalez A. The language planning situation in the Philippines. J Multilingual Multicultural Develop 1998 Sep;19(5):487-525. [doi: 10.1080/01434639808666365]
19. Kincaid J, Fishburne RP, Rogers RL, Chissom BS. Derivation Of New Readability Formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) For Navy Enlisted Personnel. Orlando: Institute for Simulation and Training, University of Central Florida; 1975.
20. Umbach PD. Web surveys: best practices. New Directions Institutional Res 2004;2004(121):23-38. [doi: 10.1002/ir.98]
21. Van Selm M, Jankowski NW. Conducting online surveys. Qual Quant 2006 Jun;40(3):435-456. [doi: 10.1007/s11135-005-8081-8]
22. King DB, O'Rourke N, DeLongis A. Social media recruitment and online data collection: a beginner's guide and best practices for accessing low-prevalence and hard-to-reach populations. Canadian Psychology/Psychologie canadienne 2014;55(4):240-249. [doi: 10.1037/a0038087]
23. National Demographic and Health Survey 2017. The World Bank. 2018. URL: https://microdata.worldbank.org/index.php/catalog/3220 [accessed 2021-11-17]
24. Lippman SA, Donini A, Díaz J, Chinaglia M, Reingold A, Kerrigan D. Social-environmental factors and protective sexual behavior among sex workers: the Encontros intervention in Brazil. Am J Public Health 2010 Apr 01;100 Suppl 1:S216-S223 [FREE Full text] [doi: 10.2105/AJPH.2008.147462] [Medline: 19762673]
25. Fonner VA, Kerrigan D, Mnisi Z, Ketende S, Kennedy CE, Baral S. Social cohesion, social participation, and HIV related risk among female sex workers in Swaziland. PLoS One 2014 Jan 31;9(1):e87527 [FREE Full text] [doi: 10.1371/journal.pone.0087527] [Medline: 24498125]
26. Grover E, Grosso A, Ketende S, Kennedy C, Fonner V, Adams D, et al. Social cohesion, social participation and HIV testing among men who have sex with men in Swaziland. AIDS Care 2016;28(6):795-804. [doi: 10.1080/09540121.2015.1131971] [Medline: 26824888]
27. Haggerty J, Lévesque JF, Santor D, Burge F, Beaulieu C, Bouharaoui F, et al. Accessibility from the patient perspective: comparison of primary healthcare evaluation instruments. Healthc Policy 2011 Dec;7(Spec Issue):94-107 [FREE Full text] [Medline: 23205038]
28. Davis C, Sloan M, Macmaster S, Hughes L. The International AIDS Questionnaire—English Version (IAQ-E). J HIV/AIDS Prevent Child Youth 2007 May 15;7(2):29-42. [doi: 10.1300/j499v07n02_03]

29. Restar AJ, Kuhns L, Reisner SL, Ogunbajo A, Garofalo R, Mimiaga MJ. Acceptability of antiretroviral pre-exposure prophylaxis from a cohort of sexually experienced young transgender women in two U.S. Cities. AIDS Behav 2018 Nov;22(11):3649-3657 [FREE Full text] [doi: 10.1007/s10461-018-2127-0] [Medline: 29713838]
30. Garofalo R, Johnson AK, Kuhns LM, Cotten C, Joseph H, Margolis A. Life skills: evaluation of a theory-driven behavioral HIV prevention intervention for young transgender women. J Urban Health 2012 Jun;89(3):419-431 [FREE Full text] [doi: 10.1007/s11524-011-9638-6] [Medline: 22223033]
31. O'brien RM. A caution regarding rules of thumb for variance inflation factors. Qual Quant 2007 Mar 13;41(5):673-690. [doi: 10.1007/s11135-006-9018-6]
32. Statacorp Stata MP 15.1 free download full version for windows. StataCorp. URL: https://www.getintopces.com/statacorp-stata-mp-15-1-free-download/ [accessed 2021-11-17]
33. Zablotska I, Grulich AE, Phanuphak N, Anand T, Janyam S, Poonkasetwattana M, et al. PrEP implementation in the Asia-Pacific region: opportunities, implementation and barriers. J Int AIDS Soc 2016 Oct 18;19(7(Suppl 6)):21119 [FREE Full text] [doi: 10.7448/IAS.19.7.21119] [Medline: 27760688]
34. Hatcher AM, Turan JM, Leslie HH, Kanya LW, Kwena Z, Johnson MO, et al. Predictors of linkage to care following community-based HIV counseling and testing in rural Kenya. AIDS Behav 2012 Jul;16(5):1295-1307 [FREE Full text] [doi: 10.1007/s10461-011-0065-1] [Medline: 22020756]
35. Jones D, Cook R, Rodriguez A, Waldrop-Valverde D. Personal HIV knowledge, appointment adherence and HIV outcomes. AIDS Behav 2013 Jan;17(1):242-249 [FREE Full text] [doi: 10.1007/s10461-012-0367-y] [Medline: 23143751]
36. Smith DK, Maier E, Betts J, Gray S, Kolodziejski B, Hoover KW. What community-based HIV prevention organizations say about their role in biomedical HIV prevention. AIDS Educ Prev 2016 Oct;28(5):426-439. [doi: 10.1521/aeap.2016.28.5.426] [Medline: 27710082]
37. Adia AC, Restar AJ, Lee CJ, Payawal MP, Quilantang MI, Nazareno J, et al. Sword and Shield: perceptions of law in empowering and protecting HIV-positive men who have sex with men in Manila, Philippines. Glob Public Health 2020 Jan;15(1):52-63. [doi: 10.1080/17441692.2019.1622762] [Medline: 31134838]
38. Calabrese SK, Krakower DS, Mayer KH. Integrating HIV Preexposure Prophylaxis (PrEP) into routine preventive health care to avoid exacerbating disparities. Am J Public Health 2017 Dec;107(12):1883-1889. [doi: 10.2105/ajph.2017.304061] [Medline: 29048955]
39. Khanna AS, Schumm P, Schneider JA. Facebook network structure and awareness of preexposure prophylaxis among young men who have sex with men. Ann Epidemiol 2017 Mar;27(3):176-180 [FREE Full text] [doi: 10.1016/j.annepidem.2016.11.006] [Medline: 28003117]
40. Dubov A, Galbo P, Altice FL, Fraenkel L. Stigma and shame experiences by MSM who take PrEP for HIV prevention: a qualitative study. Am J Mens Health 2018 Nov;12(6):1843-1854 [FREE Full text] [doi: 10.1177/1557988318797437] [Medline: 30160195]
41. Eaton LA, Kalichman SC, Price D, Finneran S, Allen A, Maksut J. Stigma and conspiracy beliefs related to Pre-exposure Prophylaxis (PrEP) and interest in using PrEP among Black and White men and transgender women who have sex with men. AIDS Behav 2017 May;21(5):1236-1246 [FREE Full text] [doi: 10.1007/s10461-017-1690-0] [Medline: 28108878]

Abbreviations

aOR: adjusted odds ratio
cis-MSM: cisgender men who have sex with men
LGBT: lesbian, gay, bisexual, and transgender
NCR: National Capital Region
NIH: National Institutes of Health
OR: odds ratio
PrEP: pre-exposure prophylaxis

Edited by R Kukafka; submitted 09.09.20; peer-reviewed by P Seekaew, T Winder; comments to author 11.10.20; revised version received 14.10.20; accepted 23.09.21; published 07.01.22.

Please cite as:
Restar A, Surace A, Adia A, Goedel W, Ogunbajo A, Jin H, Edeza A, Hernandez L, Cu-Uvin S, Operario D
Characterizing Awareness of Pre-Exposure Prophylaxis for HIV Prevention in Manila and Cebu, Philippines: Web-Based Survey of Filipino Cisgender Men Who Have Sex With Men
J Med Internet Res 2022;24(1):e24126
URL: https://www.jmir.org/2022/1/e24126
doi:10.2196/24126
PMID:34994705

Original Paper

Identifying Electronic Nicotine Delivery System Brands and Flavors on Instagram: Natural Language Processing Analysis

Rob Chew[1], MSc; Michael Wenger[1], BSc; Jamie Guillory[2], PhD; James Nonnemaker[2], PhD; Annice Kim[2], PhD

[1]Center for Data Science, RTI International, Research Triangle Park, NC, United States

[2]Center for Health Analytics, Media, and Policy, RTI International, Research Triangle Park, NC, United States

Corresponding Author:
Rob Chew, MSc
Center for Data Science
RTI International
3040 East Cornwallis Road
Research Triangle Park, NC, 27709
United States
Phone: 1 919 541 5823
Email: rchew@rti.org

Abstract

Background: Electronic nicotine delivery system (ENDS) brands, such as JUUL, used social media as a key component of their marketing strategy, which led to massive sales growth from 2015 to 2018. During this time, ENDS use rapidly increased among youths and young adults, with flavored products being particularly popular among these groups.

Objective: The aim of our study is to develop a named entity recognition (NER) model to identify potential emerging vaping brands and flavors from Instagram post text. NER is a natural language processing task for identifying specific types of words (entities) in text based on the characteristics of the entity and surrounding words.

Methods: NER models were trained on a labeled data set of 2272 Instagram posts coded for ENDS brands and flavors. We compared three types of NER models—conditional random fields, a residual convolutional neural network, and a fine-tuned distilled bidirectional encoder representations from transformers (FTDB) network—to identify brands and flavors in Instagram posts with key model outcomes of precision, recall, and F1 scores. We used data from Nielsen scanner sales and Wikipedia to create benchmark dictionaries to determine whether brands from established ENDS brand and flavor lists were mentioned in the Instagram posts in our sample. To prevent overfitting, we performed 5-fold cross-validation and reported the mean and SD of the model validation metrics across the folds.

Results: For brands, the residual convolutional neural network exhibited the highest mean precision (0.797, SD 0.084), and the FTDB exhibited the highest mean recall (0.869, SD 0.103). For flavors, the FTDB exhibited both the highest mean precision (0.860, SD 0.055) and recall (0.801, SD 0.091). All NER models outperformed the benchmark brand and flavor dictionary look-ups on mean precision, recall, and F1. Comparing between the benchmark brand lists, the larger Wikipedia list outperformed the Nielsen list in both precision and recall.

Conclusions: Our findings suggest that NER models correctly identified ENDS brands and flavors in Instagram posts at rates competitive with, or better than, others in the published literature. Brands identified during manual annotation showed little overlap with those in Nielsen scanner data, suggesting that NER models may capture emerging brands with limited sales and distribution. NER models address the challenges of manual brand identification and can be used to support future infodemiology and infoveillance studies. Brands identified on social media should be cross-validated with Nielsen and other data sources to differentiate emerging brands that have become established from those with limited sales and distribution.

(J Med Internet Res 2022;24(1):e30257) doi:10.2196/30257

KEYWORDS

named entity recognition; ENDS; social media; brands; flavors

Introduction

Background

Social media platforms provide opportunities for brands to market products to users and potential users of tobacco products [1-3]. JUUL was one of the first electronic nicotine delivery system (ENDS) brands to engage in extensive social media marketing on Instagram, Facebook, and Twitter starting in 2015, with the marketing prominently featuring sweet and fruit-flavored products (eg, mango) [4]. JUUL, to some extent, took the tobacco control community by surprise. From the company's inception, JUUL sales grew rapidly, with the company contributing substantially to the 97% growth of the ENDS marketplace between 2015 and 2018 [5,6].

Social media marketing of ENDS products has become widespread, a study conducted by the Campaign for Tobacco-Free Kids and Netnografica identifying >100 social media marketing campaigns conducted by tobacco industry giants via paid social media influencers, which elicited 8.8 billion views in the United States and 25 billion views worldwide [1]. Unsurprisingly, more than half of the youths said that they had recently seen vaping advertisements and that social media was the primary avenue of exposure [2].

Social media marketing of ENDS products coincides with the increasing prevalence of ENDS use among youths and young adults [7-11]. In 2011, current ENDS use (ie, past 30-day use) was 1.5% and 4.9% among high schoolers and middle schoolers, respectively [7], whereas, in 2019, ENDS use was 27.5% and 10.5% among high school and middle schoolers [8]. In 2013 and 2014, current ENDS use was 12.5% among young adults aged 18 to 24 years [9], whereas in 2019, ENDS use was 24.5% among the same age group [10]. Current ENDS use and JUUL use are also higher among young adults (under 24 years) than use among older adults (aged ≥25 years) [11,12]. A recent longitudinal cohort study also showed that marketing ENDS products (via web-based advertising broadcast, radio or internet radio, retail advertising, and billboards) to youths who never used ENDS products predicted ENDS initiation in young adulthood [13], establishing a link between ENDS marketing and use.

ENDS are available in a myriad of flavors (eg, fruit, dessert, mint, and menthol) that appeal to youths and young adults [14,15]. Indeed, 22% of youth ENDS users said that they used ENDS as "they are available in flavors, such as mint, candy, fruit, or chocolate," and 69% of current youth ENDS users report the use of flavored ENDS [15]. More than three-quarters of young adult ENDS users said that the first ENDS product they used was flavored [14].

The connection between social media marketing and increasing rates of ENDS use, particularly flavored ENDS, among youths and young adults underscores the importance of being able to identify emerging brands and flavors to inform tobacco control research and policy. Previous studies have typically identified brands on the web using manual searches or web scraping. For example, Zhu et al [16] identified 466 brands by searching using multiple search engines and then checking websites. Hsu et al [17], in 2016 to 2017, searched the same brand websites as identified by Zhu et al [16] and found that only 288 of the 466 brands were still active and identified 145 new brands. O'Brien et al [18] also used a manual search strategy of multiple data sources to identify tobacco product brands. These studies suggest that vaping brands are continually emerging, and a more scalable, automated method is needed to identify emerging vaping brands and flavors to inform public health surveillance.

Several recent studies have used computational models to characterize ENDS brands and flavors in vaping-related social media posts. Vandewater et al [19] identified posts from 4 brands and then used text mining to differentiate brand posts (ie, to predict which of 4 brands posted a message). Measures included length of post (eg, Blu had the shortest posts) and words used (eg, Blu and NJOY used lifestyle words, and Logic and Metro focused on product purchase, device, and use). This study showed that computational text mining could help identify and predict linguistic patterns in posts by brand. Xie et al [20] developed named entity recognition (NER) models for several entity types using text from an e-cigarette forum. NER is a natural language processing task that identifies certain words (entities) within text based on the characteristics of the entity and its surrounding words. The models were trained on a labeled data set in which specific entities were manually labeled. e-Liquid flavors were the most difficult type of entity for models to consistently identify, with F1 scores ranging from 0.0 to 0.786 across models. These studies demonstrate the potential benefit of using computational methods to analyze large volumes of social media data at scale to identify and predict patterns more quickly than traditional manual methods. To date, computational methods have not been used to identify emerging vaping brands, and limited work has demonstrated the benefit of computational methods to identify flavor mentions on social media [20].

This Study

In this study, we develop NER models to identify potential emerging vaping brands and flavors on Instagram. In our adaptation, the entities of interest are brands and flavors, and the training data comprises manually labeled brands and flavors in the Instagram post text. A benefit of using NER for brand and flavor identification is that a well-performing model will not only identify whether a post contains brand or flavor mentions but can also identify what the brand or flavor is and where in the post the brand or flavor mention occurs. Quantifying how often certain brands and flavors are mentioned can not only help in understanding the growth in vaping marketing discussions generally but may also help identify new emerging brands on social media in a systematic way.

Methods

Data Collection

Our data set included 2272 Instagram posts that were manually coded (Table 1). We used Brandwatch, a commercial social media monitoring platform, to extract all Instagram posts that appeared in the query from April 22 to April 23, 2020, with mentions related to vaping (eg, #vape, #ecigs, #vapelife, #vapecommunity) and excluding mentions of cannabis and cannabidiol (eg, #cannabis, #cbd, #420), for use in manual

coding (see Multimedia Appendix 1 for the full query). We coded the first 2272 posts from this period.

Table 1. Statistics of annotated data.

Data item	Frequency
Posts	2272
Tokens	79,401
Unique tokens	11,102
Brand mentions	1235
Flavor mentions	506

Data Annotation

We trained 2 coders to manually label posts to identify mentions of ENDS brands and flavors (see example post in Figure 1). Coders were instructed to exclude posts from coding if the post (1) did not mention a vaping brand, (2) focused on cannabis or cannabidiol products, or (3) was in a language other than English. For the purposes of brand identification and coding, we defined a brand as an ENDS product that was manufactured and sold under a specified name, which we distinguished from specified product lines (ie, specific groups of products sold under a single brand name). For example, Puff Bar is a brand name that sells several product lines, such as Puff Bar, Puff XXL, and Puff Plus. Coders also coded posts for mentions of flavors, with our definition including both common flavors that describe fruit and dessert (eg, raspberry and crème brulee) and flavors with ambiguous flavor names (eg, tropical and blue). Coders also used data on brands and flavors from other data sources (eg, Nielsen scanner data and search engine queries) to inform the coding of brands.

Coders were provided with a codebook that included the above definitions for ENDS brands and flavors with example brand and flavor names. During a preliminary training session, we showed coders a series of example Instagram posts where they were asked to identify brands and flavors in the text of posts and received feedback from a third adjudicator. To facilitate coders' annotation of the Instagram post text, we loaded Instagram post text and links to Instagram posts into a custom web-based user interface that coders accessed with a personalized username and password. Coders used this interface to annotate ENDS brands and flavors in Instagram post text by highlighting the relevant post text and using a dropdown menu to identify the highlighted text as *brand* or *flavor*. Coders were instructed to open the provided Instagram post link if additional contextual information was needed to annotate post text with ENDS brands and flavors (see Figure 2 for an example of annotated post text). During the training period, coders double coded sets of 50 test posts, and interrater reliability was assessed using Cohen κ. Once we achieved interrater reliability of Cohen $\kappa>0.70$ (brands: 0.95; flavors: 0.74), coders single coded the final sample of 2272 posts, with each coder single coding approximately half of the posts.

To prepare the data for modeling, we segmented the posts into tokens, delineated by special characters such as white space or punctuation, using the tokenization required for each model [21,22]. Hashtags were removed from the post text, and emojis were replaced by their Unicode common locale data repository text short names. The inside–outside (IO) encoding scheme was used for its simplicity and as it performed comparably with the other encoding schemes tested, including the IO-beginning and beginning-inside-last-outside-unit schemes. Figure 2 provides an example of the IO encoding scheme applied to an ENDS Instagram post; items classified as O represent text that is outside of the labeled text (brand and flavor).

Figure 1. Example of a vaping post on Instagram.

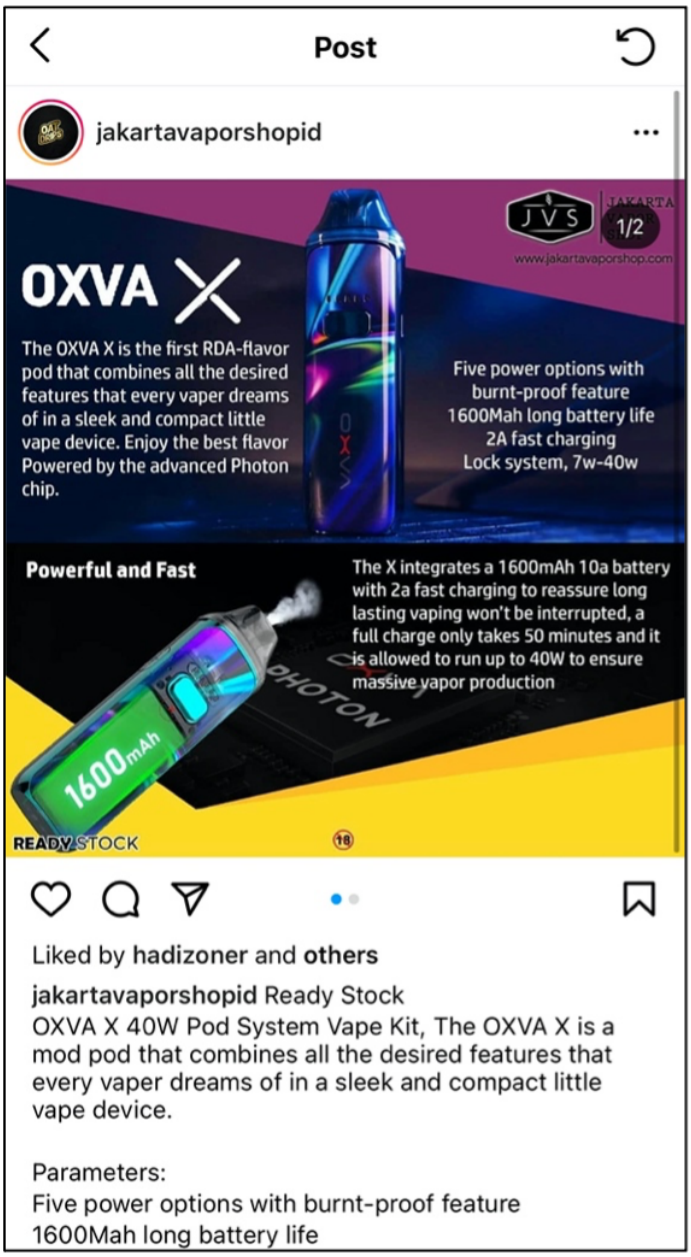
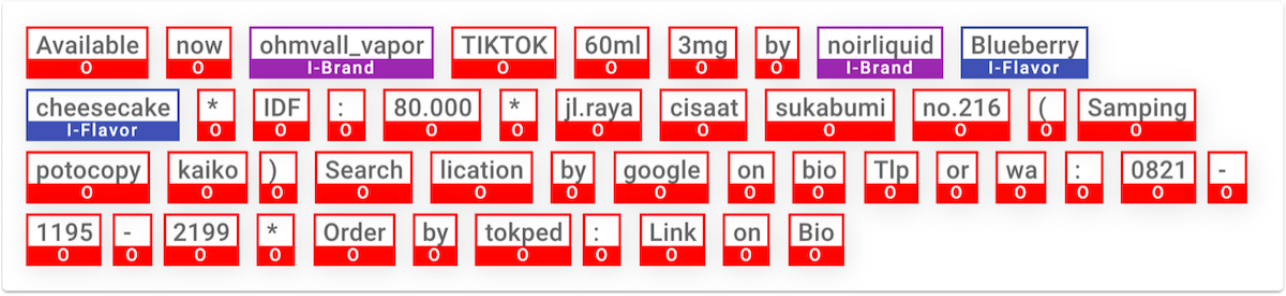


Figure 2. Example of annotated post text after inside–outside encoding was applied.



NER Models and Statistical Analysis

NER Models

Conditional Random Fields

Conditional random fields (CRFs) [23] are graphical models commonly used for NER. Although NER can be performed as a traditional classification task, where attributes of an observation are used to infer a categorical designation and each observation is treated independently, graphical models such as CRF acknowledge that text and other structured data often exhibit strong conditional dependence between neighboring observations (ie, neighboring words). This class of models predicts named entities based on an entire sentence rather than

making isolated, individual predictions for every word that are not informed by local contextual information.

CRFs model the dependencies between words as an undirected graph between labels and word attributes. These attributes were encoded by the research team as feature functions, which encapsulate the characteristics of a word that may be useful for discriminating between entity types. Textbox 1 summarizes the feature functions used as input variables for the CRF model that were calculated for each token in the training set. Feature functions can also include the characteristics of neighboring words, both before and after the current token. This set of functions represents common feature functions that have been used in other NER brand extraction models [24,25] and NER models in the vaping domain [20].

Specifically, we used a linear chain CRF, which can be thought of as a sequential extension of logistic regression [26]. To learn the model weights, we used the limited-memory Broyden–Fletcher–Goldfarb–Shannon (BFGS) constrained optimization method for gradient descent [27] with an elastic net regularization [28] condition added to the loss function. The CRF model was developed using the sklearn-crfsuite library, a Python wrapper of the popular CRFsuite software (version 0.12; Okazaki) [29].

Textbox 1. Conditional random field feature functions.

Conditional random field feature functions

- Current token is lower cased
- Current token is upper cased
- Current token is title cased
- Current token is a digit
- Prefix of current token
- Suffix of current token
- Previous token is lower cased
- Previous token is upper cased
- Previous token is title cased
- Next token is lower cased
- Next token is upper cased
- Next token is title cased

Residual Convolutional Neural Network

As an alternative to CRFs, we trained a residual convolutional neural network (RCNN) using the spaCy Python library [30]. Similar to the development of feature functions in the CRF, we created the (1) lower cased version of the token, (2) token prefix, (3) token suffix, and (4) token shape to extract potentially noteworthy features for each token in the training set. To incorporate these components into a numerical representation, each of the components was individually hashed using a Bloom filter, and the hashes were combined to create a probabilistically distinct vector [31]. In addition to this strategy being memory efficient, it also handles out-of-vocabulary tokens better than traditional word embedding approaches such as word2vec [32] as it incorporates subword information. This embedding was fed forward to a maxout layer [33] to learn a piecewise linear activation function over the inputs.

To add the context of neighboring words, the current token embedding was concatenated to the embeddings from neighboring tokens, fed to another maxout layer, and run through several residual connections [34]. A final step to consolidate the state across words was performed, which was then fed to a softmax layer to return predicted probabilities across entity types. The implementation for this study was trained from scratch using randomly initialized weights. Dropout [35] was used for regularization to help prevent overfitting, and a compounding minibatch strategy was used to increase the batch size throughout the iteration [36]. The Adam optimization algorithm [37] was used to learn the weights over 50 iterations.

Fine-tuned Distilled Bidirectional Encoder Representations From Transformers

For the final NER model, we used transfer learning to fine-tuned distilled bidirectional encoder representations from transformers (FTDB) model to our ENDS brand and flavor NER tasks using the HuggingFace Transformers library [38] in Python. Transfer learning [39] is a machine learning framework in which the objective is to use a model trained in one source domain or task to develop a model in a related target domain or task. When implemented effectively, transfer learning can provide useful representations learned in the source domain to help accelerate learning and require less labeled data in the target domain. We performed a method of transfer learning used in deep learning models called fine-tuning [40]. We started with a network model that has been trained in a source domain on a specific task, or set of tasks, and used these weights to initialize a model that is updated for a task in the target domain (eg, learning to identify ENDS brands and flavors in Instagram post text). Traditionally, model weights are initialized by assuming little knowledge about what the appropriate starting values for the weights should be. For example, random weight initialization was used to train

the RCNN model in this study. Starting with pretrained weights rather than random initializations allowed the model to initiate with components that benefit the new target task and are often challenging to learn anew with limited training data.

DistilBERT [22] is a deep neural network derived from the bidirectional encoder representations from transformers (BERT) architecture [41]. BERT is a transformer model [42], a type of neural network that takes a sequence of tokens as input and produces a contextualized vector representation of each token as its output, primarily using a mechanism called attention. BERT uses multiheaded attention layers to dynamically assign a weight to every pair of words in a sequence, where the weight indicates how much attention the model should pay to the first token when computing the representation of the second token. The multiple attention heads within each layer are trained in parallel and can each potentially capture different word–word relations. For example, Manning et al [43] showed that attention heads in BERT learned foundational linguistic characteristics such as syntactic grammatical relationships and anaphoric coreference, despite not being trained for those objectives.

BERT is pretrained on two supervised learning tasks in which labels are already encoded in the text and do not require additional manual annotation: (1) masked language modeling, which replaces random tokens in a sequence with a special mask symbol and training a model to predict the premasked token from the surrounding context, and (2) next sentence prediction, which trains a model to predict whether the next sentence is probable, given the current sentence. BERT was trained on BooksCorpus (800 million words) [44] and English Wikipedia (2500 million words), allowing it to learn masked language modeling and next sentence prediction on a massive amount of written English text. Learning these tasks on such a large corpus allowed fine-tuned BERT models to reach state-of-the-art performance on a variety of natural language understanding tasks such as question answering, semantic textual similarity, and sentiment analysis [41]. However, fine-tuning and making new predictions with BERT is still relatively resource intensive because of the number of model parameters (110 million).

DistilBERT is a compact model trained to mimic BERT that is 40% smaller and 60% faster when creating new predictions while performing similarly [22]. The model compression was achieved by using a method called knowledge distillation [45,46], in which a *student* model (distilled BERT [distilBERT]) is trained to reproduce the behavior of a larger *teacher* model (BERT), in this case, by training distilBERT to predict the BERT's output class probabilities. Our NER model fine-tunes the distilBERT weights by training it on the task of ENDS brand and flavor token recognition using labeled data from Instagram posts. During training, we used a batch size of 16, a weight decay of 0.01, and 500 warm-up steps over the course of 3 epochs.

Benchmarks

Overview

As a comparison to the NER models, we created benchmarks to check if brands from an established ENDS brand list are mentioned in the Instagram posts. This approach helped us to better understand coverage gaps between brands mentioned on social media versus established brand lists and helped to assess whether NER modeling aids in emerging brand detection. Although not created for cataloging ENDS brands and flavors on social media, these benchmarks represent data sets that researchers may naturally be inclined to use for ENDS brand and flavor identification on social media, because of their popularity or availability, or as a reference when developing their own databases.

To determine if a post contained a brand from an existing list, we performed a dictionary look-up of the brand or flavor names within the post text. Any mention of the brand or flavor name within the post text was flagged as a potential mention, favoring higher recall over precision. In addition, we compared the top-selling brands in the Nielsen Retail scanner data set with the most mentioned brands in our hand-labeled Instagram data to better understand differences in brand composition.

ENDS Brands and Flavors From Nielsen Retail Scanner Data Set

The Nielsen scanner data set for ENDS brands (N=66) is a unique list of brands derived from any sales of ENDS products in large food and convenience stores in the United States from March 22, 2020, to May 16, 2020, a span that overlaps with the dates of the Instagram posts in our data set. These data also include information on ENDS flavors (N=219) sold over the same period.

ENDS Brands From Wikipedia Data

The Wikipedia data set for ENDS brands (N=249) is a publicly available, crowdsourced list of electronic cigarettes and vaping liquid brands [47]. Although less informative as a ground-truth source of ENDS sales data, the Wikipedia list is larger, providing a potentially more comprehensive and diverse list of ENDS brands than the Nielsen scanner data set, which focuses on top-selling brands. Wikipedia entries and lists have been used and repurposed in various studies [48,49]. The version of the list used in this study was accessed on March 16, 2020.

Validation

Overview

To prevent overfitting, we used a 5-fold cross-validation methodology. Cross-validation requires creating models on different, mutually exclusive partitions of training and validation observations. For 5-fold cross-validation, 5 different models (using the same model type) were created and assessed on 5 nonoverlapping validation sets. By noting differences in model performance across the runs, we can gain a more realistic range of results for how a model will generalize on new samples, assuming that the samples are drawn from the same underlying distribution as the training data.

We used the standard evaluation metrics to assess different aspects of the predictions. All metrics are bounded between 0 and 1, and, all else being equal, a higher value indicates a better performing model.

Precision

Precision is the fraction of true positives out of all observations that are predicted to be positive.

Precision = true positive / (true positive + false positive) **(1)**

Recall

Recall is the fraction of true positives detected out of all positive labeled examples.

>Recall = true positive / (true positive + false negative) **(2)**

F1

This summary metric balances the need to find positive examples with the need to reduce the number of false positives.

F1 = 2 / ((1 / recall) + (1 / precision)) **(3)**

Results

Brands

Table 2 summarizes the NER and benchmark dictionary look-up approaches across a 5-fold cross-validation scheme. For brands, the mean precision was higher than the mean recall across all methods except for the FTDB, ranging from 0.082 to 0.797 for mean precision and 0.022 to 0.869 for mean recall. The RCNN exhibited the highest mean precision (0.797, SD 0.084) for brands, and the FTDB exhibited the highest mean recall (0.869, SD 0.103). However, despite the FTDB exhibiting the highest mean recall, it also had the largest SD in recall scores across the validation folds for brands (0.103). The FTDB had the highest mean F1 score for brands (0.815, SD 0.097), substantially outperforming the approach with the second-highest F1 score, the CRF (0.551, SD 0.037).

Table 2. Named entity recognition and benchmark validation for identifying brand and flavor mentions in vaping Instagram posts.

Approach	Value, mean (SD)		
	Precision	Recall	F1 score
Nielsen scanner d ata match			
Brand	0.082 (0.061)	0.022 (0.015)	0.034 (0.024)
Flavor	0.507 (0.102)	0.417 (0.088)	0.446 (0.053)
Wikipedia d ata match			
Brand	0.224 (0.039)	0.102 (0.025)	0.140 (0.031)
Flavor	N/A[a]	N/A	N/A
CRF [b]			
Brand	0.670 (0.080)	0.472 (0.044)	0.551 (0.037)
Flavor	0.695 (0.201)	0.496 (0.058)	0.573 (0.109)
RCNN [c]			
Brand	0.797 (0.084)	0.405 (0.036)	0.535 (0.032)
Flavor	0.753 (0.232)	0.454 (0.082)	0.549 (0.109)
FTDB [d]			
Brand	0.768 (0.095)	0.869 (0.103)	0.815 (0.097)
Flavor	0.860 (0.055)	0.801 (0.091)	0.828 (0.069)

[a]N/A: not applicable.

[b]CRF: conditional random field.

[c]RCNN: residual convolutional neural network.

[d]FTDB: fine-tuned distilled bidirectional encoder representations from transformers.

Flavors

For flavors, the mean precision was uniformly higher than the mean recall across all approaches. The flavor mean precision values ranged from 0.507 to 0.860, and the mean recall values ranged from 0.417 to 0.801. The FTDB exhibited the highest mean precision (0.860, SD 0.05), recall (0.801, SD 0.091), and F1 score (0.828, SD 0.069) for the flavor NER task. It also had the lowest SD of precision values across validation folds (0.055) when compared with the other approaches.

Benchmark Comparisons

Overall, the machine learning NER models outperformed the benchmark matching approaches across the mean precision, recall, and F1 metrics, with comparable SDs. The major exception is the precision of the flavor mentions, which have greater variation in the RCNN and CRF NER models (RCNN: SD 0.232; CRF: SD 0.201) than the Nielsen scanner data matching (SD 0.102). However, the FTDB model exhibited the lowest SD of precision values for flavor mentions (SD 0.055), despite it also being a machine learning approach.

Comparing benchmarks, the larger Wikipedia data set performed better on brands for both precision (Nielsen: mean 0.082, SD 0.061; Wikipedia: mean 0.224, SD 0.039) and recall (Nielsen: mean 0.022, SD 0.015; Wikipedia: mean 0.102, SD 0.025). An ENDS product flavor list was not available from Wikipedia for comparison with the scanner data.

Table 3 compares the brands with the most mentions in our annotated Instagram data, using a 30% random test set of posts, with the top-selling brands in the Nielsen scanner data. Besides Puff Bar, there is little overlap between the top mentioned brands on Instagram and the top-selling brands identified through scanner data during the study period. Established brands comprising the largest market share of ENDS product sales (JUUL and Vuse) are not well-represented in our sample, which comprises emerging brands (Smok and GeekVape) and smaller niche brands (Majestea, a brand specializing in tea-flavored vape e-liquids and nicotine salts).

Table 3. Top brands in the Nielsen scanner and annotated Instagram data set by market share and brand mention share.

Brand	Share, n (%)
Nielsen scanner (market; US $)	
JUUL	$330,397,506 (58.43)
Vuse	$119,245,199 (21.09)
Puff Bar	$38,852,586 (6.87)
NJOY	$27,522,703 (4.87)
Blu	$16,128,231 (2.85)
Logic	$7,045,441 (1.25)
Pop	$6,963,975 (1.23)
Bidi Stick	$6,693,510 (1.18)
ES	$2,982,509 (0.53)
Jak	$2,292,140 (0.41)
Instagram (mention; n=392)	
Smok	31 (7.9)
GeekVape	24 (6.1)
Puff Bar	16 (4.1)
This is Salts	16 (4.1)
Adore eLiquid	13 (3.3)
Chief of Vapes	9 (2.3)
Majestea	9 (2.3)
Villain Vapors	9 (2.3)
Conspiracy	6 (1.5)
Orgnx	6 (1.5)

Discussion

Principal Findings

This study extends the ENDS literature by using computational methods to identify brand and flavor mentions on social media. Our findings suggest that NER models can improve upon a naïve approach of searching for known brands and flavors in noisy Instagram post text, both in terms of precision and recall. A notable finding is the superior performance of the FTDB model over all the approaches measured in this study. Although large increases in model performance are not uncommon when starting from pretrained models for computer vision and natural language processing tasks (see similarly large increases in performance on the WNUT17 NER data set in the *Comparison With Prior Works* section), the sizable leap in mean F1 scores between the FTDB (brands: 0.815; flavors: 0.828) when compared with the second best-performing CRF (brands: 0.551; flavors: 0.573) suggests that fine-tuning pretrained models for NER tasks on social media is worth the added complexity. In particular, the FTDB model far outperformed the other NER models in mean recall. Although all the NER models had reasonable precision (ie, terms predicted as brands or flavors were usually correct), the CRF and RCNN predicted a sizably smaller proportion of the true labeled brands and flavors in the validation sets than the FTDB.

The performance of the Wikipedia and Nielsen scanner brand dictionary look-up implies that a larger, more comprehensive list may improve the identification of known brands on Instagram. However, the many false positives may also suggest that some brand names are too generic when matched in isolation (eg, Carbon, Square, Epic, and Zoom). This may also indicate differences in coverage between popular brands in

terms of sales and those being actively marketed and discussed on social media. The low level of overlap between the top-selling brands in the Nielsen scanner data and the most mentioned brands in our data set suggests that differing brand coverage may be a contributing factor. Given that current popular ENDS brands such as JUUL have been widely discussed on social media before the firm's meteoric market growth [4], understanding ENDS brand engagement on social media may be an early indicator of emerging brands before they surface in sales data. Brands identified on social media should continually be cross-validated with other data sources, such as Nielsen, to better understand whether these models identify (1) emerging brands that will eventually become more established brands, (2) brands that will come and go, or (3) smaller-scale brands that are sold only on the web or in vapor shops.

Comparison With Prior Work

Most published works on brand and flavor–named entity models come from research teams affiliated with large retail or e-commerce companies (eBay, Amazon, and Walmart) that use NER models to parse product titles or descriptions. These studies generally report precision and recall of >0.80 for brands and flavors [24,25,50,51]. However, despite the popularity of brand and flavor NER models on retail sites, there is considerably less published research on brand and flavor NER models for social media data. To our knowledge, the closest brand example is the WNUT17 data set [52] designed for benchmarking models to detect emerging named entities on social media (Twitter, Reddit, YouTube, and StackExchange). Although this data set does not contain a *brand* entity type, it does contain entities for *corporation* and *product*. Across 7 research teams, the original paper reported F1 scores on this task that ranged from 0.253 to 0.402 for detecting any instances of entity names and from 0.263 to 0.419 when only unique entity names were counted. More recently, newer models using fine-tuned transformer-based architectures have accounted for F1 scores in the range of 0.580 [53] to 0.600 [54]. This decline in NER performance on social media when compared with product titles and descriptions highlights the diversity of linguistic patterns and the use of unique syntactic conventions that make working with social media data challenging. Importantly, the authors note that *corporation* was generally a difficult class for systems to predict and speculate that this may be partly because of confusion between corporate and product entity types. For NER models with a flavor entity type, Xie et al [20] identified e-liquid flavors from posts on an e-cigarette forum. They reported an F1 score of 0.786, the lowest of all the 9 entity types they assessed. Although the reported metrics in this study are higher than those in the study by Derczynski et al [52], the test set for flavors in their study was comprised of only 13 observations, making it challenging to compare and draw strong generalization claims. Although direct comparisons with our models are complicated by differences in the sample composition, entity definitions, and social media platform, our best cross-validation results for brands (F1 score: mean 0.815, SD 0.097) outperform previous NER results on social media for entity types conceptually close to brands [52], and our results for ENDS flavors (F1 score: mean 0.828, SD 0.069) also outperform those previously reported in the literature [20].

Implications

To predict the ENDS brands that may emerge and become popular among youths, there is a need to monitor social media platforms for ENDS advertising and conversations. An important component of this monitoring is to identify ENDS brands to be tracked and monitored on social media and other data sources (eg, scanner data) to better understand their marketing, appeal, and potential for growth to help the tobacco control community identify the next JUUL. Relying solely on traditional data sources such as Nielsen sales data or Kantar advertising expenditure data may not be adequate for helping to identify emerging brands. Indeed, new brands entering the marketplace may start by promoting their products on social media through organic posts (paid advertising of tobacco products is not allowed on social media platforms, including Facebook, Instagram, and Twitter) and selling products on the web as the barrier to entry is low. These brands would not appear in other traditional sales data sources; Nielsen data only captures brick-and-mortar retail sales and not web-based sales data. These data are also unlikely to show up in traditional advertising data sources, such as Kantar data, which only captures paid advertising and not organic social media posts. Brands that appear in sales and advertising data are likely to be more established brands spending money on advertising, distribution, and sales in retail stores.

Limitations

Our study has several limitations. First, labeling brands is not always a straightforward task for human coders, especially when there is a need to distinguish between brands, products, and product lines. In addition, labeling flavors is complicated in the ENDS product space because of the presence of both *characterizing flavors* (flavors that describe existing food, beverages, or spices, such as grape, apple, and cinnamon) and *concept flavors* (ambiguous flavor names not tied to existing food, beverages, or spices, such as Bayou Blast and Midnight Madness) [55]. Difficulty in distinguishing between brand or corporation and product entities on social media [52] and difficulty in consistently identifying e-liquid flavors on social media [20] have been documented in the literature. Although we established sufficiently high intercoder reliability, our coders went through several rounds of calibration as there were disagreements about how to label certain mentions within posts. When there is ambiguity in how to label categories, or if labeling certain types of entities is inherently difficult, the accuracy of predictions for supervised machine learning models may decrease [56]. Second, most NER model metrics published in the literature report exact matches to the original labels, down to the character level. There may be cases where the model identified a ground-truth brand label; however, the predicted span was slightly offset, or only a portion of the full brand name was captured (eg, the model identified *BANG Bars*, but the annotation was *BANG Bars XL*). If partial matches are useful, the NER model evaluation metrics may understate the value of the model predictions. Although some works acknowledge this issue and incorporate partial match evaluation metrics into their reporting [57], this practice is not yet common in the literature. Third, as this study focused on a sample of Instagram posts, findings may not generalize beyond Instagram users to the

general ENDS user population. Furthermore, although we theorized that the language used to describe ENDS brands and flavors in our study time frame is not characteristically different from other periods in the recent past, this hypothesis has not been tested and may benefit from further research. Finally, increasing the number of training examples could help improve model performance. Unlike more structured text examples for brand and flavor extraction, such as product titles, ENDS-related Instagram posts vary greatly stylistically, from marketing posts using concentrated bursts of product information to consumer posts that only mention vaping off hand. Although using transfer learning somewhat mitigates the need for more training data, having a larger corpus of example posts that better captures this diversity should improve model performance. In addition, methods for data augmentation [58] may also help improve performance.

Future Work

Our results suggest several natural extensions for future work. First, as Instagram is a visual social medium, brand and flavor information is often conveyed in post images and text. Future work could use a multi-view learning approach [59] to incorporate information from both images and posts to improve model performance [60]. Longer-term NER models can be combined with entity linking [61] to update a database of established and emerging vaping brands and flavors. Entity linking, the process of reconciling extracted entities into a canonical entry, is often useful to combine with NER models because of the various ways brands and flavors can be mentioned in social media text (eg, extracted brand mentions of *puffbar*, *puff bar*, *puff bars*, and *p u f f b a r* should be associated with the canonical brand name *PuffBar*). This approach borrows the strengths of both the NER modeling and benchmark dictionary approaches outlined in this study. Although only brand and flavor NER models were examined in this work, extending to other named entities can help answer more nuanced research questions relevant to the tobacco control community. For example, NER results for relevant sentence subjects (brand or flavor mentions) could be linked to NER results for sentence objects relevant to public health outcomes, such as health claims [62] or adverse events [63,64]. NER can also be useful as a building block for more advanced downstream tasks, such as entity-based sentiment analysis [65] to better understand the sentiment associated with a given named entity (eg, the sentiment associated with certain brands or flavors). Finally, NER models could be used to support future infodemiology or infoveillance studies. Infodemiology [66,67] is an emerging science at the intersection of consumer and public health informatics that involves using publicly available data that can be identified, analyzed, and stored to provide important public health insights (eg, using web-based search query data to predict influenza outbreaks). For example, NER models could be used to identify and track ENDS brands or flavors over time to better understand whether their growth, or change in composition, is associated with related population health outcomes.

Conclusions

This study demonstrates that NER models can be used to correctly identify ENDS brands and flavors in Instagram posts, with models showing favorable performance to previous research using NER to identify entities similar to brands and ENDS flavors in social media posts. The brands identified in our models had little overlap with brands from Nielsen scanner data, suggesting that these may be emerging brands. NER models may be a helpful approach for identifying emerging brands on social media before they appear in other data sources that capture more established brands. NER models address the challenges of manual brand identification being time consuming and difficult without pre-existing brand dictionaries, making them attractive for new brand identification. Brands identified on social media should continually be validated against other data sources, such as Nielsen scanner data, to help differentiate emerging brands that have become established from those with limited sales and distribution.

Acknowledgments

Funding for this study was provided by a contract with the Florida Department of Health (contract number: COTGC).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Query for identifying electronic nicotine delivery system Instagram posts.
[DOCX File , 39 KB - jmir_v24i1e30257_app1.docx]

References

1. New investigation exposes how tobacco companies market cigarettes on social media in the U.S. and around the world. Campaign for Tobacco-Free Kids. 2018. URL: https://www.tobaccofreekids.org/press-releases/2018_08_27_ftc [accessed 2021-12-17]
2. Cho YJ, Thrasher JF, Reid JL, Hitchman S, Hammond D. Youth self-reported exposure to and perceptions of vaping advertisements: findings from the 2017 International Tobacco Control Youth Tobacco and Vaping Survey. Prev Med 2019 Sep;126:105775 [FREE Full text] [doi: 10.1016/j.ypmed.2019.105775] [Medline: 31323286]

3. Myers M, Muggli M, Henigan D. Request for investigative and enforcement action to stop deceptive advertising online. Tobacco Free Kids. 2018. URL: https://www.tobaccofreekids.org/assets/content/press_office/2018/2018_08_ftc_petition.pdf [accessed 2021-12-17]
4. Jackler R, Chau C, Getachew B, Whitcomb M, Lee-Heidenreich J, Bhatt A, et al. JUUL advertising over its first three years on the market. Stanford Research into the Impact of Tobacco Advertising. 2019. URL: https://tobacco-img.stanford.edu/wp-content/uploads/2021/07/21231836/JUUL_Marketing_Stanford.pdf [accessed 2021-12-17]
5. Craver R. Juul expands top U.S. e-cig market share; traditional cigarettes volume continues to slip. Winston–Salem Journal. 2018. URL: https://journalnow.com/business/juul-expands-top-u-s-e-cig-market-share-traditional/article_9bdfd55c-68b5-5c08-aeb8-edb4a616ca9e.html [accessed 2021-12-18]
6. LaVito A. Popular e-cigarette Juul's sales have surged almost 800 percent over the past year. CNBC. 2018. URL: https://www.cnbc.com/2018/07/02/juul-e-cigarette-sales-have-surged-over-the-past-year.html [accessed 2018-12-05]
7. Cullen KA, Ambrose BK, Gentzke AS, Apelberg BJ, Jamal A, King BA. Notes from the field: use of electronic cigarettes and any tobacco product among middle and high school students - United States, 2011-2018. MMWR Morb Mortal Wkly Rep 2018 Nov 16;67(45):1276-1277 [FREE Full text] [doi: 10.15585/mmwr.mm6745a5] [Medline: 30439875]
8. Cullen KA, Gentzke AS, Sawdey MD, Chang JT, Anic GM, Wang TW, et al. e-cigarette use among youth in the United States, 2019. J Am Med Assoc 2019 Dec 03;322(21):2095-2103 [FREE Full text] [doi: 10.1001/jama.2019.18387] [Medline: 31688912]
9. Stanton CA, Sharma E, Edwards KC, Halenar MJ, Taylor KA, Kasza KA, et al. Longitudinal transitions of exclusive and polytobacco electronic nicotine delivery systems (ENDS) use among youth, young adults and adults in the USA: findings from the PATH Study Waves 1-3 (2013-2016). Tob Control 2020 May 22;29(Suppl 3):147-154 [FREE Full text] [doi: 10.1136/tobaccocontrol-2019-055574] [Medline: 32321848]
10. Cornelius ME, Wang TW, Jamal A, Loretan CG, Neff LJ. Tobacco product use among adults - United States, 2019. MMWR Morb Mortal Wkly Rep 2020 Nov 20;69(46):1736-1742 [FREE Full text] [doi: 10.15585/mmwr.mm6946a4] [Medline: 33211681]
11. Prakash S, Hatcher C, Shiffman S. Prevalence of ENDS and JUUL use, by Smoking Status, in national samples of young adults and older adults in the U.S. Am J Health Behav 2021 May 01;45(3):402-418. [doi: 10.5993/ajhb.45.3.2]
12. Vallone DM, Bennett M, Xiao H, Pitzer L, Hair EC. Prevalence and correlates of JUUL use among a national sample of youth and young adults. Tob Control 2019 Nov 29;28(6):603-609. [doi: 10.1136/tobaccocontrol-2018-054693] [Medline: 30377241]
13. Loukas A, Paddock EM, Li X, Harrell MB, Pasch KE, Perry CL. Electronic nicotine delivery systems marketing and initiation among youth and young adults. Pediatrics 2019 Sep 26;144(3):e20183601 [FREE Full text] [doi: 10.1542/peds.2018-3601] [Medline: 31451608]
14. Rostron BL, Cheng Y, Gardner LD, Ambrose BK. Prevalence and reasons for use of flavored cigars and ENDS among US youth and adults: estimates from Wave 4 of the PATH Study, 2016-2017. Am J Health Behav 2020 Jan 01;44(1):76-81. [doi: 10.5993/ajhb.44.1.8]
15. Wang TW, Gentzke AS, Creamer MR, Cullen KA, Holder-Hayes E, Sawdey MD, et al. Tobacco product use and associated factors among middle and high school students - United States, 2019. MMWR Surveill Summ 2019 Nov 06;68(12):1-22 [FREE Full text] [doi: 10.15585/mmwr.ss6812a1] [Medline: 31805035]
16. Zhu S, Sun JY, Bonnevie E, Cummins SE, Gamst A, Yin L, et al. Four hundred and sixty brands of e-cigarettes and counting: implications for product regulation. Tob Control 2014 Jul 16;23 Suppl 3(suppl 3):3-9 [FREE Full text] [doi: 10.1136/tobaccocontrol-2014-051670] [Medline: 24935895]
17. Hsu G, Sun J, Zhu S. Evolution of electronic cigarette brands from 2013-2014 to 2016-2017: analysis of brand websites. J Med Internet Res 2018 Mar 12;20(3):e80 [FREE Full text] [doi: 10.2196/jmir.8550] [Medline: 29530840]
18. O'Brien EK, Hoffman L, Navarro MA, Ganz O. Social media use by leading US e-cigarette, cigarette, smokeless tobacco, cigar and hookah brands. Tob Control 2020 Dec 26;29(e1):87-97. [doi: 10.1136/tobaccocontrol-2019-055406] [Medline: 32217772]
19. Vandewater EA, Clendennen SL, Hébert ET, Bigman G, Jackson CD, Wilkinson AV, et al. Whose post is it? Predicting e-cigarette brand from social media posts. Tob Regul Sci 2018 Mar 01;4(2):30-43. [doi: 10.18001/trs.4.2.3]
20. Xie J, Liu X, Zeng D. Mining e-cigarette adverse events in social media using Bi-LSTM recurrent neural network with word embedding representation. J Am Med Inform Assoc 2018 Jan 01;25(1):72-80 [FREE Full text] [doi: 10.1093/jamia/ocx045] [Medline: 28505280]
21. Honnibal M, Ines M, Sofie V, Adriane B. spaCy: industrial-strength natural language processing in Python. spaCy. URL: https://spacy.io/ [accessed 2021-01-21]
22. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. arXiv. 2019. URL: https://arxiv.org/abs/1910.01108 [accessed 2021-12-17]
23. Lafferty J, McCallum A, Pereira F. Conditional random fields: probabilistic models for segmenting and labeling sequence data. ACM Digital Library. URL: http://portal.acm.org/citation.cfm?id=655813 [accessed 2021-12-17]
24. More A. Attribute extraction from product titles in eCommerce. arXiv. 2016. URL: https://arxiv.org/abs/1608.04670 [accessed 2021-12-17]

25. Putthividhya D, Hu J. Bootstrapped named entity recognition for product attribute extraction. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July 2011; Edinburgh, Scotland, UK p. 1557-1567 URL: https://aclanthology.org/D11-1144/
26. Sutton C, McCallum A. An introduction to conditional random fields for relational learning. In: Introduction to Statistical Relational Learning. Cambridge, MA: MIT Press; 2007:2-128.
27. Zhu C, Byrd RH, Lu P, Nocedal J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. ACM Trans Math Softw 1997 Dec;23(4):550-560. [doi: 10.1145/279232.279236]
28. Zou H, Hastie T. Regularization and variable selection via the elastic net. J Royal Statistical Soc B 2005 Apr;67(2):301-320. [doi: 10.1111/j.1467-9868.2005.00503.x]
29. Okazaki N. CRFsuite: a fast implementation of Conditional Random Fields (CRFs). GitHub. 2007. URL: https://github.com/chokkan/crfsuite [accessed 2021-12-17]
30. Honnibal M, Montani I. spaCy 2: natural language understanding with Bloom embeddings, convolutional neural networks and incremental parsing. spaCy. URL: https://spacy.io/ [accessed 2021-01-21]
31. Serra J, Karatzoglou A. Getting deep recommenders fit: bloom embeddings for sparse binary input/output networks. In: Proceedings of the Eleventh ACM Conference on Recommender Systems. 2017 Presented at: RecSys '17: Eleventh ACM Conference on Recommender Systems; August 27 - 31, 2017; Como Italy p. 279-287. [doi: 10.1145/3109859.3109876]
32. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv. 2013. URL: https://arxiv.org/abs/1301.3781 [accessed 2021-12-17]
33. Goodfellow I, Warde-Farley D, Mirza M, Courville A, Bengio Y. Maxout networks. Proc Mach Learn Res 2013;28(3):1319-1327 [FREE Full text]
34. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016 Presented at: IEEE Conference on Computer Vision and Pattern Recognition (CVPR); June 27-30, 2016; Las Vegas, NV, USA. [doi: 10.1109/cvpr.2016.90]
35. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 2014;15(56):1929-1958 [FREE Full text]
36. Smith S, Kindermans P, Ying C, Le Q. Don't decay the learning rate, increase the batch size. arXiv. 2017. URL: https://arxiv.org/abs/1711.00489 [accessed 2021-12-17]
37. Kingma D, Ba J. Adam: a method for stochastic optimization. ResearchGate. 2014. URL: https://www.researchgate.net/publication/269935079_Adam_A_Method_for_Stochastic_Optimization [accessed 2021-12-17]
38. Wolf T, Debut L, Sanh V, Chaumond J, Delangue C, Moi A, et al. HuggingFace's transformers: state-of-the-art natural language processing. arXiv. 2019. URL: https://arxiv.org/abs/1910.03771 [accessed 2021-12-17]
39. Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010 Oct;22(10):1345-1359. [doi: 10.1109/tkde.2009.191]
40. Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: Proceedings of the International Conference on Artificial Neural Networks. 2018 Presented at: International Conference on Artificial Neural Networks; October 4-7, 2018; Rhodes, Greece p. 270-279. [doi: 10.1007/978-3-030-01424-7_27]
41. Devlin J, Chang M, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. arXiv. 2018. URL: https://arxiv.org/abs/1810.04805 [accessed 2021-12-17]
42. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A, et al. Attention is all you need. NeurIPS Proceedings. 2017. URL: https://papers.nips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html [accessed 2021-12-17]
43. Manning CD, Clark K, Hewitt J, Khandelwal U, Levy O. Emergent linguistic structure in artificial neural networks trained by self-supervision. Proc Natl Acad Sci U S A 2020 Dec 01;117(48):30046-30054 [FREE Full text] [doi: 10.1073/pnas.1907367117] [Medline: 32493748]
44. Zhu Y, Kiros R, Zemel R, Salakhutdinov R, Urtasun R, Torralba A, et al. Aligning books and movies: Towards story-like visual explanations by watching movies and reading books. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2015 Presented at: IEEE International Conference on Computer Vision (ICCV); Dec. 7-13, 2015; Santiago, Chile. [doi: 10.1109/iccv.2015.11]
45. Buciluǎ C, Caruana R, Niculescu-Mizil A. Model compression. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2006 Presented at: KDD06: The 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 20 - 23, 2006; Philadelphia PA USA p. 535-541. [doi: 10.1145/1150402.1150464]
46. Hinton G, Vinyals O, Dean J. Distilling the knowledge in a neural network. arXiv. 2015. URL: https://arxiv.org/abs/1503.02531 [accessed 2021-12-17]
47. List of electronic cigarette and e-cigarette liquid brands. Wikipedia. URL: https://en.wikipedia.org/wiki/List_of_electronic_cigarette_and_e-cigarette_liquid_brands [accessed 2021-03-16]
48. Samoilenko A, Yasseri T. The distorted mirror of Wikipedia: a quantitative analysis of Wikipedia coverage of academics. EPJ Data Sci 2014 Jan 22;3(1):1-11. [doi: 10.1140/epjds20]
49. Roll U, Mittermeier JC, Diaz GI, Novosolov M, Feldman A, Itescu Y, et al. Using Wikipedia page views to explore the cultural importance of global reptiles. Biol Conserv 2016 Dec;204:42-50. [doi: 10.1016/j.biocon.2016.03.037]

50. Zheng G, Mukherjee S, Dong X, Li F. OpenTag: open attribute value extraction from product profiles. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018 Presented at: KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 19 - 23, 2018; London United Kingdom p. 1049-1058. [doi: 10.1145/3219819.3219839]
51. Karamanolakis G, Ma J, Dong X. Txtract: taxonomy-aware knowledge extraction for thousands of product categories. arXiv. 2020. URL: https://arxiv.org/abs/2004.13852 [accessed 2021-12-17]
52. Derczynski L, Nichols E, van EM, Limsopatham N. Results of the WNUT2017 shared task on novel and emerging entity recognition. In: Proceedings of the 3rd Workshop on Noisy User-Generated Text. 2017 Presented at: 3rd Workshop on Noisy User-generated Text; September 2017; Copenhagen, Denmark p. 140-147. [doi: 10.18653/v1/w17-4418]
53. Ushio A, Camacho-Collados J. T-NER: an all-round python library for transformer-based named entity recognition. In: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations. 2021 Presented at: 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations; Apri, 2021; Online p. 53-62. [doi: 10.18653/v1/2021.eacl-demos.7]
54. Wang X, Jiang Y, Bach N, Wang T, Huang Z, Huang F, et al. Improving named entity recognition by external context retrieving and cooperative learning. arXiv. 2021. URL: https://arxiv.org/abs/2105.03654 [accessed 2021-12-18]
55. Carpenter CM, Wayne GF, Pauly JL, Koh HK, Connolly GN. New cigarette brands with flavors that appeal to youth: tobacco marketing strategies. Health Aff (Millwood) 2005 Nov;24(6):1601-1610. [doi: 10.1377/hlthaff.24.6.1601] [Medline: 16284034]
56. Frenay B, Verleysen M. Classification in the presence of label noise: a survey. IEEE Trans Neural Netw Learn Syst 2014 May;25(5):845-869. [doi: 10.1109/tnnls.2013.2292894]
57. Segura-Bedmar I, Martínez P, Herrero-Zazo M. SemEval-2013 Task 9: extraction of drug-drug interactions from biomedical texts (DDIExtraction 2013). In: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013). 2014 Oct Presented at: Seventh International Workshop on Semantic Evaluation (SemEval 2013); June 14-15, 2013; Atlanta, Georgia p. 341-350 URL: https://e-archivo.uc3m.es/bitstream/handle/10016/20455/semeval_SEMEVAL_2013.pdf?sequence=3
58. Wei J, Zou K. EDA: easy data augmentation techniques for boosting performance on text classification tasks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019 Presented at: Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); November 2019; Hong Kong, China p. 6382-6388. [doi: 10.18653/v1/d19-1670]
59. Li Y, Yang M, Zhang Z. A survey of multi-view representation learning. IEEE Trans Knowl Data Eng 2019 Oct 1;31(10):1863-1883. [doi: 10.1109/tkde.2018.2872063]
60. Moon S, Neves L, Carvalho V. Multimodal named entity recognition for short social media posts. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). 2018 Presented at: Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers); June 2018; New Orleans, Louisiana p. 852-860. [doi: 10.18653/v1/n18-1078]
61. Shen W, Wang J, Han J. Entity linking with a knowledge base: issues, techniques, and solutions. IEEE Trans Knowl Data Eng 2015 Feb 1;27(2):443-460. [doi: 10.1109/tkde.2014.2327028]
62. Yuan S, Yu B. HClaimE: a tool for identifying health claims in health news headlines. Inf Process Manag 2019 Jul;56(4):1220-1233. [doi: 10.1016/j.ipm.2019.03.001]
63. Chen X, Deldossi M, Aboukhamis R, Faviez C, Karapetiantz B, Guenegou-Arnoux A, et al. Mining adverse drug reactions in social media with named entity recognition and semantic methods. ResearchGate. 2017. URL: https://tinyurl.com/2fky3yeb [accessed 2021-12-18]
64. MacKinlay A, Aamer H, Yepes A. Detection of adverse drug reactions using medical named entities on Twitter. AMIA Annu Symp Proc 2017;2017:1215-1224 [FREE Full text] [Medline: 29854190]
65. Saif H, He Y, Fernandez M, Alani H. Contextual semantics for sentiment analysis of Twitter. Inf Process Manag 2016 Jan;52(1):5-19. [doi: 10.1016/j.ipm.2015.01.005]
66. Eysenbach G. Infodemiology and infoveillance: framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. J Med Internet Res 2009 Mar 27;11(1):e11 [FREE Full text] [doi: 10.2196/jmir.1157] [Medline: 19329408]
67. Eysenbach G. Infodemiology and infoveillance tracking online health information and cyberbehavior for public health. Am J Prev Med 2011 May;40(5 Suppl 2):154-158. [doi: 10.1016/j.amepre.2011.02.006] [Medline: 21521589]

Abbreviations

BERT: bidirectional encoder representations from transformers
BFGS: Broyden–Fletcher–Goldfarb–Shannon
CRF: conditional random field

distilBERT: distilled bidirectional encoder representations from transformers
ENDS: electronic nicotine delivery system
FTDB: fine-tuned distilled bidirectional encoder representations from transformers
IO: inside–outside
NER: named entity recognition
RCNN: residual convolutional neural network

Edited by R Kukafka; submitted 09.05.21; peer-reviewed by Z Xie, S Doan, A Mavragani; comments to author 17.07.21; revised version received 01.11.21; accepted 21.11.21; published 18.01.22.

Please cite as:
Chew R, Wenger M, Guillory J, Nonnemaker J, Kim A
Identifying Electronic Nicotine Delivery System Brands and Flavors on Instagram: Natural Language Processing Analysis
J Med Internet Res 2022;24(1):e30257
URL: https://www.jmir.org/2022/1/e30257
doi:10.2196/30257
PMID:35040793

Original Paper

A Novel Virtual Reality Assessment of Functional Cognition: Validation Study

Lilla Alexandra Porffy[1], BSc, MSc; Mitul A Mehta[1], PhD; Joel Patchitt[1,2], BSc; Celia Boussebaa[1], BSc; Jack Brett[3], BSc; Teresa D'Oliveira[1], PhD; Elias Mouchlianitis[4*], PhD; Sukhi S Shergill[1,5,6*], MBBS, PhD

[1]Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom

[2]Trafford Centre for Medical Research, University of Sussex, Brighton, United Kingdom

[3]Faculty of Media and Communications, Bournemouth University, Bournemouth, United Kingdom

[4]School of Psychology, University of East London, London, United Kingdom

[5]Kent and Medway Medical School, Canterbuy, United Kingdom

[6]Kent and Medway National Heath Service and Social Care Partnership Trust, Gillingham, United Kingdom

[*]these authors contributed equally

Corresponding Author:
Lilla Alexandra Porffy, BSc, MSc
Institute of Psychiatry, Psychology & Neuroscience, King's College London
16 De Crespigny Park
London, SE5 8AF
United Kingdom
Phone: 44 07738244217
Email: lilla.a.porffy@kcl.ac.uk

Abstract

Background: Cognitive deficits are present in several neuropsychiatric disorders, including Alzheimer disease, schizophrenia, and depression. Assessments used to measure cognition in these disorders are time-consuming, burdensome, and have low ecological validity. To address these limitations, we developed a novel virtual reality shopping task—VStore.

Objective: This study aims to establish the construct validity of VStore in relation to the established computerized cognitive battery, Cogstate, and explore its sensitivity to age-related cognitive decline.

Methods: A total of 142 healthy volunteers aged 20-79 years participated in the study. The main VStore outcomes included verbal recall of 12 grocery items, time to collect items, time to select items on a self-checkout machine, time to make the payment, time to order coffee, and total completion time. Construct validity was examined through a series of backward elimination regression models to establish which Cogstate tasks, measuring attention, processing speed, verbal and visual learning, working memory, executive function, and paired associate learning, in addition to age and technological familiarity, best predicted VStore performance. In addition, 2 ridge regression and 2 logistic regression models supplemented with receiver operating characteristic curves were built, with VStore outcomes in the first model and Cogstate outcomes in the second model entered as predictors of age and age cohorts, respectively.

Results: Overall VStore performance, as indexed by the total time spent completing the task, was best explained by Cogstate tasks measuring attention, working memory, paired associate learning, and age and technological familiarity, accounting for 47% of the variance. In addition, with λ=5.16, the ridge regression model selected 5 parameters for VStore when predicting age (mean squared error 185.80, SE 19.34), and with λ=9.49 for Cogstate, the model selected all 8 tasks (mean squared error 226.80, SE 23.48). Finally, VStore was found to be highly sensitive (87%) and specific (91.7%) to age cohorts, with 94.6% of the area under the receiver operating characteristic curve.

Conclusions: Our findings suggest that VStore is a promising assessment that engages standard cognitive domains and is sensitive to age-related cognitive decline.

(J Med Internet Res 2022;24(1):e27641) doi:10.2196/27641

KEYWORDS

virtual reality; virtual reality assessment; cognition; functional cognition; functional capacity; neuropsychological testing

Introduction

Background

Cognitive dysfunction refers to deficits in intellectual functions usually described by domains such as attention, working memory, verbal and visual learning, executive function, and processing speed. Deficits in cognition are evident across a range of neuropsychiatric disorders, including Alzheimer disease (AD), schizophrenia, and depression. Although these intellectual deficits are diagnostic in AD [1], 90% of individuals with schizophrenia and depression are also affected [2,3]. This is further complicated by the observation that some cognitive decline is part of the natural aging process and is reported in one-quarter of older adults without dementia [4]. The high prevalence of cognitive dysfunction in mental and physical [5] illness, an increasingly aging population, and the lack of robust treatments suggest that the global burden of cognitive dysfunction has a substantial socioeconomic impact.

Cognitive decline has a marked effect on functional recovery and quality of life in patients with mental disorders [6-8]. In addition, it precedes and predicts functional outcomes in AD and schizophrenia [9,10] and predicts treatment response in depression [11], highlighting the urgent need to effectively target these symptoms. Unfortunately, clinical trials of cognitive enhancers have been largely disappointing [12-14]. Indeed, most compounds that demonstrated positive effects in phase 2 trials have failed in phase 3 trials. This raises several questions about the sensitivity of our cognitive assessments and the targets of these interventions.

Standard cognitive assessments are designed to evaluate changes in distinct neuropsychological domains, whereas the actual target for therapy is change in functional cognition—the ability to perform everyday routine activities [15]. Accordingly, the Food and Drug Administration has mandated the assessment of real-life functional change, alongside changes in conventional cognitive performance, as a condition for drug approval for both AD and schizophrenia [16,17]. This is particularly important as there can be a lack of concordance between changes in cognitive measures and related everyday functioning. For example, cognitive task performance only explains 20% of the variance in work-related skills in schizophrenia [18]. Although there has been an attempt to supplement cognitive assessments with self-report and reports by caregivers to assess wider functioning, these assessments lack objectivity.

A related issue is that cognitive assessments require optimal task engagement, which can be confounded by poor attention and motivation [19,20]. The gold standard cognitive measure for AD, the Alzheimer's Disease Assessment Scale–Cognitive Subscale [21], takes approximately 45 minutes to administer; the analogous scale for schizophrenia, the MATRICS Consensus Cognitive Battery [22,23], takes up to 90 minutes. The related functional capacity assessment for AD, the Clinical Dementia Rating Scale [24], takes approximately 30 minutes to complete, similar to the University of California, San Diego Performance-Based Skills Assessment [25] for schizophrenia. The ecological validity and predictive power of real-life performance of standard assessments have also been questioned [26,27].

Complex assessments that emulate everyday scenarios have been developed, including the Multiple Errands Test (MET), which measures executive function in patients with traumatic brain injury [28], and the Virtual Reality Functional Capacity Assessment Tool (VRFCAT), which measures functional skills in schizophrenia and can reliably differentiate patients from controls [29,30]. However, the MET is time-consuming and difficult to standardize with a lack of experimental control, whereas the VRFCAT lacks full ecological validity as it is completed on a computer or tablet without the immersive nature of real-life interactions.

Objectives

Recent developments in technology, specifically in virtual reality (VR), now enable us to create assessments that can replicate challenges found in everyday life while also maintaining experimental control [31]. This offers the opportunity to overcome issues associated with current assessments. In this study, we describe the development of a novel, fully immersive VR assessment, VStore, with the aim to simultaneously assess traditional cognitive domains and functional capacity. This is achieved through the creation of an ecologically valid minimarket environment with a maze-like layout. Each action within the assessment maps an embedded cognitive task (eg, recall of shopping list items measures verbal memory), and each task is assessed by performing actions that require almost identical procedures similar to shopping in real life, offering a measure of concurrent functional capacity. Moving in an immersive VR environment engages brain structures associated with spatial navigation, such as the hippocampus and entorhinal cortex [32], which are affected in early AD [33], depression [34], and schizophrenia [35]. Therefore, VStore may be more sensitive to early neurodegenerative processes than the existing assessments.

The aim of this study is 2-fold. First, we establish the cognitive domains relevant to VStore performance. More specifically, we test which cognitive processes, as measured by an existing standard cognitive battery, predict VStore performance as an initial evaluation of its construct. We achieve this by conducting a series of stepwise prediction models. Second, we explore the preliminary utility of VStore in assessing cognitive decline associated with nonpathological aging. This is achieved by testing VStore's ability to predict age both as a continuous and dichotomized outcome.

Methods

Participants

A total of 142 healthy volunteers aged 20-79 years were recruited through advertisements in college circular emails, charity newsletters, and social media. Participants were excluded if they had (1) a diagnosis of an axis 1 disorder (Diagnostic and Statistical Manual of Mental Disorders, 5th edition); (2) dependence on alcohol or illicit substances; (3) clinically significant motion sickness; (4) a pregnancy; and (5) a diagnosis of a neurological illness. Of the 142 volunteers, 38 (26.8%)

participants were excluded from the study. The reasons for exclusion were as follows: 1 participant withdrew consent, 1 could not complete VStore owing to technical issues, 1 senior participant could not complete VStore owing to fatigue, 20 participants failed either or both integrity and completion criteria for Cogstate, and 15 participants were removed owing to outlier values on one or more primary outcome measures. The demographic information for the final sample of 73.2% (104/142) of participants is presented in Table 1.

Table 1. Sample demographics.

Variable	Age group (years)						
	20-29	30-39	40-49	50-59	60-69	70-79	Total
Population, n (%)	19 (18.3)	18 (17.3)	18 (17.3)	17 (16.3)	18 (17.3)	14 (13.5)	104 (100)
Age (years), mean (SD; range)	23.6 (2.5; 20-29)	32.7 (2.3; 30-37)	45.2 (2.8; 40-49)	53.5 (3.0; 50-59)	64.2 (2.4; 61-69)	73.4 (3.3; 70-79)	48.8 (17.8; 20-79)
Gender (female), n (%)	10 (52.6)	11 (61.1)	9 (50)	8 (47.1)	10 (55.6)	6 (42.9)	54 (51.9)
IQ, mean (SD; range)	119.7 (9.5; 96-136)	121.3 (6.9; 105-131)	120.6 (6.0; 106-132)	121.7 (7.9; 109- 133)	126.3 (5.9; 113-134)	128.1 (6.3; 118-138)	122.9 (10.3; 96-138)
Education (years); mean (SD; range)	17.6 (2.4; 15-24)	19.4 (3.0; 14-27)	18.8 (1.9; 15-22)	18.0 (9.5; 11-20)	16.1 (3.4; 10-23)	18.1 (5.3; 10-25)	18.0 (3.6; 10-27)
Technological familiarity, mean (SD; range)	47.1 (6.8; 34-57)	47.2 (6.4; 32-57)	43.5 (7.9; 32-58)	43.4 (7.9; 28-57)	35.6 (8.3; 21-52)	33.4 (8.6; 21-48)	41.7 (9.7; 21-58)

Measures

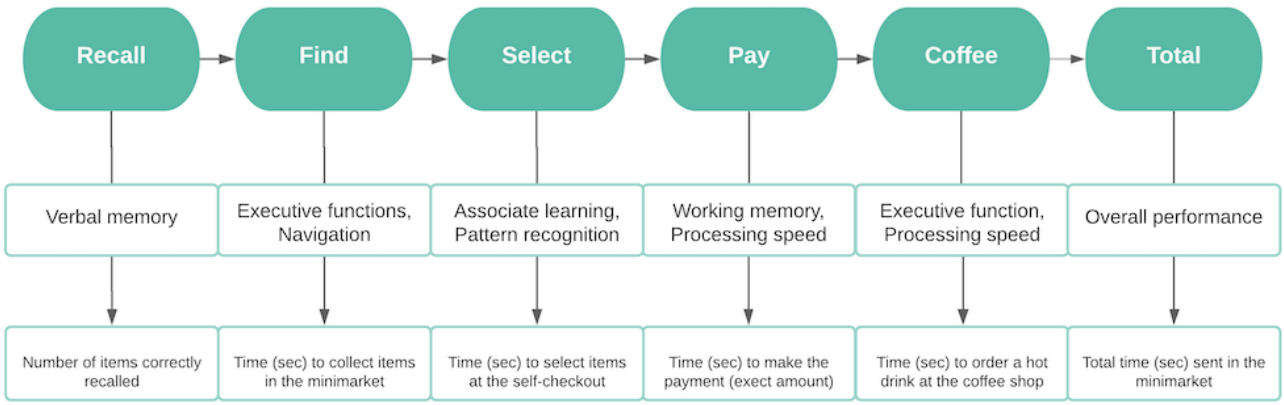
VStore

VStore was developed in collaboration with Vitae VR [36]. It takes approximately 30 minutes to complete, including orientation, instructions, practice, and assessment. Orientation and practice are set in a courtyard specifically designed for VR acclimatization (Multimedia Appendix 1).

The assessment itself is set in a minimarket environment depicting a fruit and vegetable section; 6 aisles of foodstuff, snacks, drinks, and toiletries; fridges with chilled drinks and sandwiches; and freezers with frozen meals. In addition, there are checkout and self-checkout counters and a coffee shop at the back of the minimarket. A total of 66 items, organized into 9 categories, were created to fill the shop (Multimedia Appendix 2).

At the start, participants were read out 12 items from a shopping list (Multimedia Appendix 3) by the avatar standing near the entrance. The first task of the participants was to memorize and recall as many items from this list as possible. Following recall, participants were presented with the shopping list, including all 12 items, and instructed to move around the shop and collect all items as quickly and accurately as possible. Once all the items are bagged, they are required to select and pay for them at a self-checkout machine, providing the exact amount (Multimedia Appendix 4). The task is concluded when participants order a hot drink from the coffee shop situated in the minimarket. Progression to the next task could only be achieved after successfully completing the previous task. The steps required to complete the VStore tasks are summarized in Figure 1. Multimedia Appendices 5-7 provide details on apparatus information, software information, and how movement is executed in the virtual environment, respectively.

Figure 1. Flowchart depicting the steps required to complete VStore, its corresponding cognitive domains, and outcome variables.



Cogstate

Cogstate is a computerized cognitive battery designed to assess multiple cognitive domains. It has been widely used in both healthy and clinical populations. Cogstate is simple to use, even for adults with limited computer experience, and therefore suitable for testing older adults [37]. For the purposes of this study, 8 tasks that cover key cognitive domains (Table 2 and Multimedia Appendix 8) were selected, taking approximately 30-40 minutes to complete.

Table 2. List of Cogstate tasks, corresponding cognitive domains assessed, and main outcome measures.

Code	Task name	Cognitive domain	Outcome (metric)
DET	Detection	Processing speed	Reaction time (log 10 ms)
IDN	Identification	Attention	Reaction time (log 10 ms)
OCL	One Card Learning	Visual learning	Accuracy (arcsine proportion)
ONB	One-Back	Working memory	Reaction time (log 10 ms)
TWO	Two-Back	Working memory	Accuracy (arcsine proportion)
CPAL	Continuous Paired Associate Learning	Paired associate learning	Total number of errors (N/A[a])
GMLT	Groton Maze Learning	Executive functions	Total number of errors (N/A)
ISLT	International Shopping List Task	Verbal learning	Number of correct responses (N/A)

[a]N/A: not applicable.

Wechsler Abbreviated Scale of Intelligence

The abbreviated version of the Wechsler Adult Intelligence Scale was used to establish the IQ of participants [38]. Specifically, the 2-item scale included matrix and vocabulary tests.

Technological Familiarity Questionnaire

We developed a self-report questionnaire to assess the technological familiarity of the sample population. Participants were asked 13 questions to ascertain their frequency, comfort, and ability in technology use. Higher scores indicated more technological familiarity. The internal consistency of the questionnaire was good (Cronbach α=.88). A detailed description of the Technological Familiarity Questionnaire (TFQ) is presented in Multimedia Appendix 9.

Procedures

Potential participants were prescreened over the phone. If they were deemed eligible, they were invited for a single study visit that lasted up to 2.5 hours. First, informed consent was obtained, followed by obtaining demographics, brief mental and physical health history, and the TFQ scores. Cogstate and VStore were administered in a counterbalanced fashion to mitigate any order effects. All the participants received the same shopping list. Finally, the Wechsler Abbreviated Scale of Intelligence was administered. Participants were compensated for their time and reimbursed for travel expenses. Ethical approval was granted by the Psychiatry, Nursing and Midwifery Research Ethics Committee, King's College London (LRS-16/17-4540).

Analysis

Before data analysis, VStore outcome variables measured in seconds were log-transformed to stabilize the variance. Descriptive statistics for both VStore and Cogstate outcomes are presented in Multimedia Appendices 10 and 11. As an initial overview of the relationship between Cogstate and VStore, Bonferroni-corrected Spearman ρ was calculated between the 2 assessments. These results are presented in Multimedia Appendix 12.

To establish which cognitive domains, assessed by Cogstate, best predicted VStore performance, we ran a series of backward elimination regression models implemented in the R package MASS [39]. VStore outcomes were entered as dependent variables (DVs) and all 8 Cogstate tasks were entered as independent variables (IVs). Age and technological familiarity (TFQ) were also entered as IVs, as these (but not IQ) showed a significant relationship with VStore outcomes. All IVs were standardized using the sample mean and SD to create *z* scores. Regression models were penalized for complexity using the Akaike Information Criterion (AIC) to arrive at the most parsimonious model. Additional quality checks for the final models are presented in Multimedia Appendix 13. These confirm that the assumptions of normality and homoscedasticity were met.

As an exploratory objective to examine the potential of VStore in predicting age, we used ridge regression, implemented in the R package glmnet [40], where regularization is governed by 2 parameters— α and λ. We set the penalty parameter, α, to 0 (to enforce ridge regression, where the estimated coefficients of strongly correlated predictor variables are *shrunk* toward each other). The optimal value of the strength of this penalty (λ) was determined using leave-one-out cross-validation (ie, for a given value of λ, training on N-1 participants, and testing performance on the one participant who is held-out by computing the mean squared error [MSE]). The DV was age for 104 participants. In the first model, IVs included all VStore outcomes except for total time: Recall, Find, Select, Pay, and Coffee. In the second model, IVs included all Cogstate tasks: Detection (DET), Identification (IDN), One Card Learning (OCL), One-Back (ONB), Two-Back (TWO), Continuous Paired Associate Learning (CPAL), Groton Maze Learning (GMLT), and the International Shopping List Task (ISLT). Both models were repeated with technological familiarity (TFQ) included as an

additional IV to indicate whether VStore was confounded by technological familiarity. Finally, to further probe VStore's sensitivity in predicting age cohorts, we took the top and bottom 20% of the sample population based on age and ran 2 logistic regression models to generate 2 overlying receiver operating characteristic curves—one for VStore and one for Cogstate. The bottom fraction of the sample included 23.1% (24/104) of participants aged 20-30 years, whereas the top fraction included 22.1% (23/104) of participants aged 65-79 years. Similar to the regression analyses, the age cohort (0, 1) was entered as the DV, and IVs for VStore model were Recall, Find, Select, Pay, and Coffee, whereas the IVs for the Cogstate model included DET, IDN, OCL, ONB, TWO, CPAL, GMLT, and ISLT. Youden *J* statistic was used to establish the optimal threshold for sensitivity and specificity, and model performance was compared with the DeLong test.

Results

VStore Construct

Tables 3-5 summarizes the predictors of VStore performance. The initial model included all Cogstate variables, in addition to age and technological familiarity. Backward elimination regression resulted in the removal of several of these predictors, without any substantial change in the variance explained by the models. AIC values showed a decrease from the initial to final models, arriving at a more parsimonious set of predictors for each VStore outcome.

Recalling items from VStore shopping list was predicted by verbal learning. Finding items in VStore was best explained by attention (IDN), working memory (ONB), paired associate learning (CPAL), age, and technological familiarity (TFQ). The best predictors of VStore Select were working memory (TWO), executive functions (GMLT), verbal learning (ISLT), and age. Paying for items in VStore was best explained by processing speed (DET), working memory (TWO), executive function (GMLT), verbal learning (ISLT), and technological familiarity (TFQ). Time to order a coffee was best predicted by visual (OCL) and verbal (ISLT) learning, working memory (TWO), and age. Finally, total time spent in VStore was best explained by attention (IDN), working memory (ONB), paired associate learning (CPAL), age, and technological familiarity (TFQ). For the final model, the explained variance ranged from 25% for VStore Select to 47% for VStore Total time.

Given the prominent role of technological familiarity in VStore performance, we also examined the correlations between the TFQ and Cogstate for comparison. Indeed, 6 out of 8 Cogstate tasks (DET, IDN, ONB, TWO, CPAL, and GMLT) had a significant relationship with the TFQ (Multimedia Appendix 14).

Table 3. Initial and final linear regression models examining the construct validity of VStore for Recall, Find, and Select outcomes.

DV[a] and IV[b]	Initial model					Final model				
	B (SE)	*P* value	R^2	*F* test (*df*)	AIC[c]	B (SE)	*P* value	R^2	*F* test (*df*)	AIC
Recall										
DET[d]	0.264 (0.205)	.20	N/A[e]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IDN[f]	−0.064 (0.224)	.78	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
OCL[g]	0.078 (0.190)	.68	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ONB[h]	−0.016 (0.520)	.95	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TWO[i]	0.246 (0.218)	.26	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
GMLT[j]	−0.056 (0.214)	.79	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CPAL[k]	0.209 (0.214)	.33	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ISLT[l]	0.569 (0.202)	.006	N/A	N/A	N/A	0.056 (0.014)	<.001	N/A	N/A	N/A
Age	−0.312 (0.242)	.20	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TFQ[m]	0.001 (0.228)	.95	.10	2.138 (10, 93)	123.6	N/A	N/A	.13	16.210[n] (1, 102)	111.8
Find										
DET	0.019 (0.028)	.047	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IDN	−0.061 (0.030)	.93	N/A	N/A	N/A	−0.052 (0.028)	.07	N/A	N/A	N/A
OCL	−0.002 (0.026)	.06	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ONB	0.065 (0.034)	.92	N/A	N/A	N/A	0.074 (0.030)	.03	N/A	N/A	N/A
TWO	−0.003 (0.029)	.40	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
GMLT	0.024 (0.029)	.11	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CPAL	0.046 (0.029)	.84	N/A	N/A	N/A	0.055 (0.025)	.03	N/A	N/A	N/A
ISLT	−0.006 (0.027)	.006	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Age	0.093 (0.033)	.03	N/A	N/A	N/A	0.100 (0.029)	.001	N/A	N/A	N/A
TFQ	−0.070 (0.031)	.50	.40	7.833[n] (10, 93)	−293.1	−0.068 (0.028)	.02	.42	15.830[n] (5, 98)	−301.1
Select										
DET	−0.006 (0.034)	.86	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IDN	−0.035 (0.037)	.35	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

DV[a] and IV[b]	Initial model					Final model				
	B (SE)	*P* value	R^2	*F* test (*df*)	AIC[c]	B (SE)	*P* value	R^2	*F* test (*df*)	AIC
OCL	−0.009 (0.031)	.77	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ONB	0.020 (0.041)	.64	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TWO	−0.034 (0.036)	.35	N/A	N/A	N/A	−0.043 (0.030)	.16	N/A	N/A	N/A
GMLT	0.041 (0.035)	.252	N/A	N/A	N/A	0.043 (0.031)	.16	N/A	N/A	N/A
CPAL	0.024 (0.035)	.50	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ISLT	−0.041 (0.033)	.22	N/A	N/A	N/A	−0.047 (0.030)	.12	N/A	N/A	N/A
Age	0.111 (0.040)	.007	N/A	N/A	N/A	0.108 (0.030)	.001	N/A	N/A	N/A
TFQ	0.001 (0.038)	.97	.22	3.896[n] (10, 93)	−251.6	N/A	N/A	.25	9.783[n] (4, 99)	−261.8

[a]DV: dependent variable.

[b]IV: independent variable.

[c]AIC: Akaike Information Criterion.

[d]DET: Detection (processing speed).

[e]N/A: not applicable.

[f]IDN: Identification (attention).

[g]OCL: One Card Learning (visual learning).

[h]ONB: One-Back (working memory).

[i]TWO: Two-Back (working memory).

[j]GMLT: Groton Maze Learning (executive function).

[k]CPAL: Continuous Paired Associate Learning (paired associate learning).

[l]ISLT: International Shopping List Task (Verbal learning).

[m]TFQ: Technological Familiarity Questionnaire.

[n]Significant at $P<.001$.

Table 4. Initial and final linear regression models examining the construct validity of VStore for Pay and Coffee outcomes.

DV[a] and IV[b]	Initial model					Final model				
	B (SE)	*P* value	R^2	*F* test (*df*)	AIC[c]	B (SE)	*P* value	R^2	*F* test (*df*)	AIC
Pay										
DET[d]	0.043 (0.040)	.28	N/A[e]	N/A	N/A	0.063 (0.035)	.08	N/A	N/A	N/A
IDN[f]	0.024 (0.043)	.59	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
OCL[g]	0.032 (0.037)	.38	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ONB[h]	0.034 (0.048)	.49	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TWO[i]	−0.085 (0.042)	.047	N/A	N/A	N/A	−0.085 (0.036)	.02	N/A	N/A	N/A
GMLT[j]	0.070 (0.041)	.09	N/A	N/A	N/A	0.060 (0.037)	.11	N/A	N/A	N/A
CPAL[k]	−0.025 (0.041)	.55	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ISLT[l]	−0.072 (0.039)	.07	N/A	N/A	N/A	−0.077 (0.034)	.03	N/A	N/A	N/A
Age	0.020 (0.047)	.67	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TFQ[m]	−0.106 (0.044)	.02	.34	6.409[n] (10, 93)	−218.6	−0.129 (0.035)	<.001	.36	12.630[n] (5, 98)	−225.8
Coffee										
DET	0.018 (0.048)	.72	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IDN	−0.018 (0.053)	.73	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
OCL	0.075 (0.045)	.10	N/A	N/A	N/A	0.077 (0.041)	.07	N/A	N/A	N/A
ONB	0.021 (0.059)	.72	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
TWO	−0.046 (0.051)	.37	N/A	N/A	N/A	−0.078 (0.042)	.07	N/A	N/A	N/A
GMLT	−0.003 (0.051)	.96	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CPAL	0.061 (0.051)	.23	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ISLT	−0.060 (0.048)	.21	N/A	N/A	N/A	−0.066 (0.044)	.13	N/A	N/A	N/A
Age	0.181 (0.057)	.002	N/A	N/A	N/A	0.218 (0.043)	<.001	N/A	N/A	N/A
TFQ	−0.027 (0.054)	.62	.27	4.751[n] (10, 93)	−177.0	N/A	N/A	.30	11.810[n] (4, 99)	−186.6

[a]DV: dependent variable.

[b]IV: independent variable.

[c]AIC: Akaike Information Criterion.

[d]DET: Detection (processing speed).

[e]N/A: not applicable.

[f]IDN: Identification (attention).

[g]OCL: One Card Learning (visual learning).
[h]ONB: One-Back (working memory).
[i]TWO: Two-Back (working memory).
[j]GMLT: Groton Maze Learning (executive function).
[k]CPAL: Continuous Paired Associate Learning (paired associate learning).
[l]ISLT: International Shopping List Task (Verbal learning).
[m]TFQ: Technological Familiarity Questionnaire.
[n]Significant at $P<.001$.

Table 5. Initial and final linear regression models examining the construct validity of VStore Total.

DV[a] and IV[b]	Initial model					Final model				
	B (SE)	*P* value	R^2	*F* test (*df*)	AIC[c]	B (SE)	*P* value	R^2	*F* test (*df*)	AIC
Total										
DET[d]	0.018 (0.020)	.44	N/A[e]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
IDN[f]	−0.046 (0.024)	.08	N/A	N/A	N/A	−0.039 (0.025)	.12	N/A	N/A	N/A
OCL[g]	0.003 (0.026)	.88	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
ONB[h]	0.048 (0.022)	.10	N/A	N/A	N/A	0.067 (0.026)	.01	N/A	N/A	N/A
TWO[i]	−0.015 (0.029)	.56	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
GMLT[j]	0.025 (0.025)	.32	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
CPAL[k]	0.043 (0.025)	.09	N/A	N/A	N/A	0.059 (0.021)	.007	N/A	N/A	N/A
ISLT[l]	−0.023 (0.025)	.34	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Age	0.097 (0.024)	.001	N/A	N/A	N/A	0.109 (0.025)	<.001	N/A	N/A	N/A
TFQ[m]	−0.050 (0.028)	.06	.46	9.803[n] (10, 93)	−323.3	−0.045 (0.025)	.07	.47	19.070[n] (5, 98)	−329.1

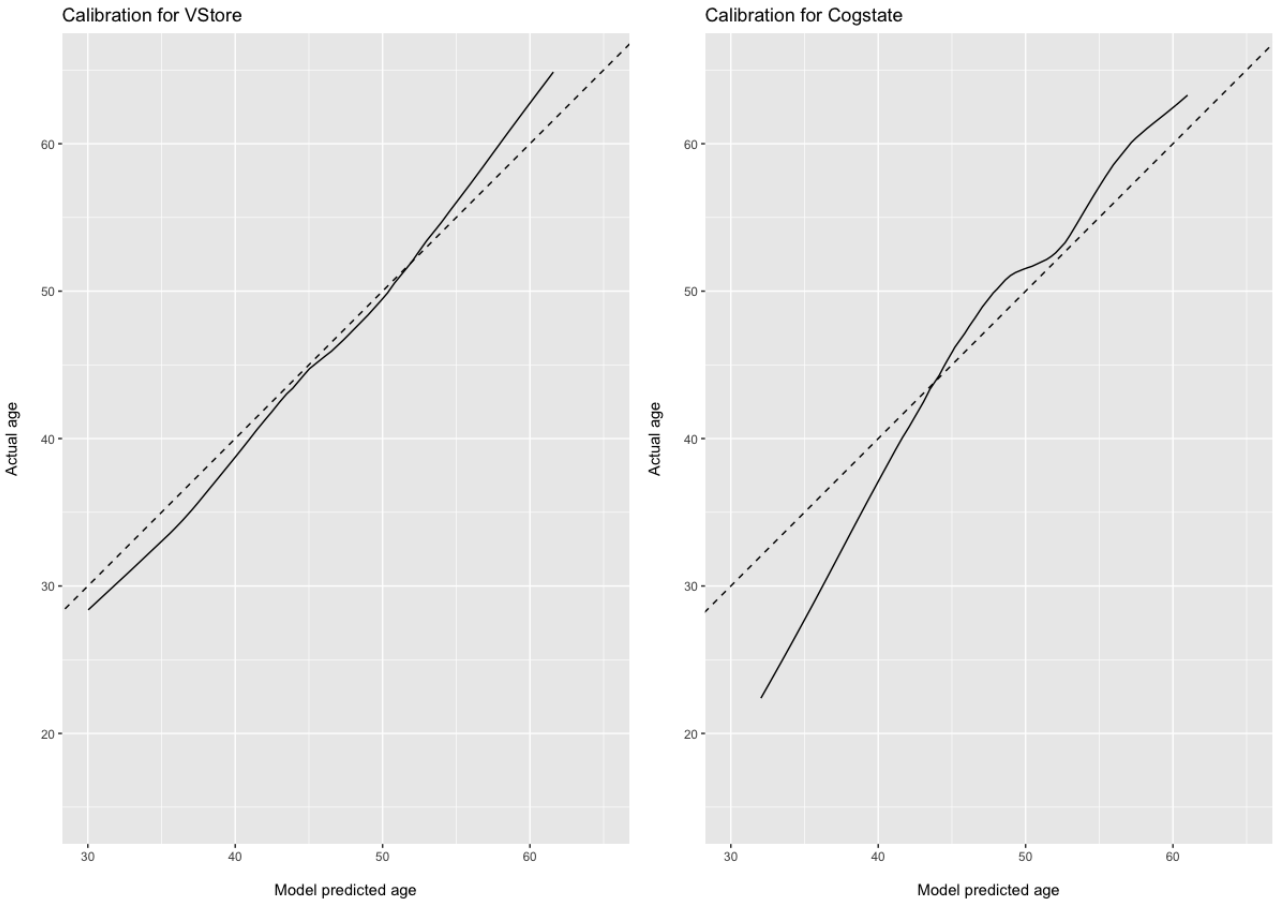
[a]DV: dependent variable.
[b]IV: independent variable.
[c]AIC: Akaike Information Criterion.
[d]DET: Detection (processing speed).
[e]N/A: not applicable.
[f]IDN: Identification (attention).
[g]OCL: One Card Learning (visual learning).
[h]ONB: One-Back (working memory).
[i]TWO: Two-Back (working memory).
[j]GMLT: Groton Maze Learning (executive function).
[k]CPAL: Continuous Paired Associate Learning (paired associate learning).
[l]ISLT: International Shopping List Task (Verbal learning).
[m]TFQ: Technological Familiarity Questionnaire.
[n]Significant at $P<.001$.

Cognitive Performance as Predictor of Age

For the DV age, we built 2 models using VStore and Cogstate outcomes as predictors (Figure 2). In the VStore model, the model fitting achieved an MSE of 185.8 (SE 19.34), selecting a total of 5 predictors and from cross-validating, which was attained from an optimal λ of 5.16 (Multimedia Appendix 15). For the Cogstate model, we found an MSE of 226.8 (SE 23.48), selecting a total of 8 predictors obtained with an optimal λ of 9.49 (Multimedia Appendix 16). We also fitted a null

(intercept-only) model that yields an MSE of 294.71, suggesting that models for both VStore and Cogstate are preferable to a model with no predictors.

Figure 2. VStore and Cogstate models predicting age. Dashed lines depict the age of the participants. The solid stroke shows the age predicted by the ridge regression models.

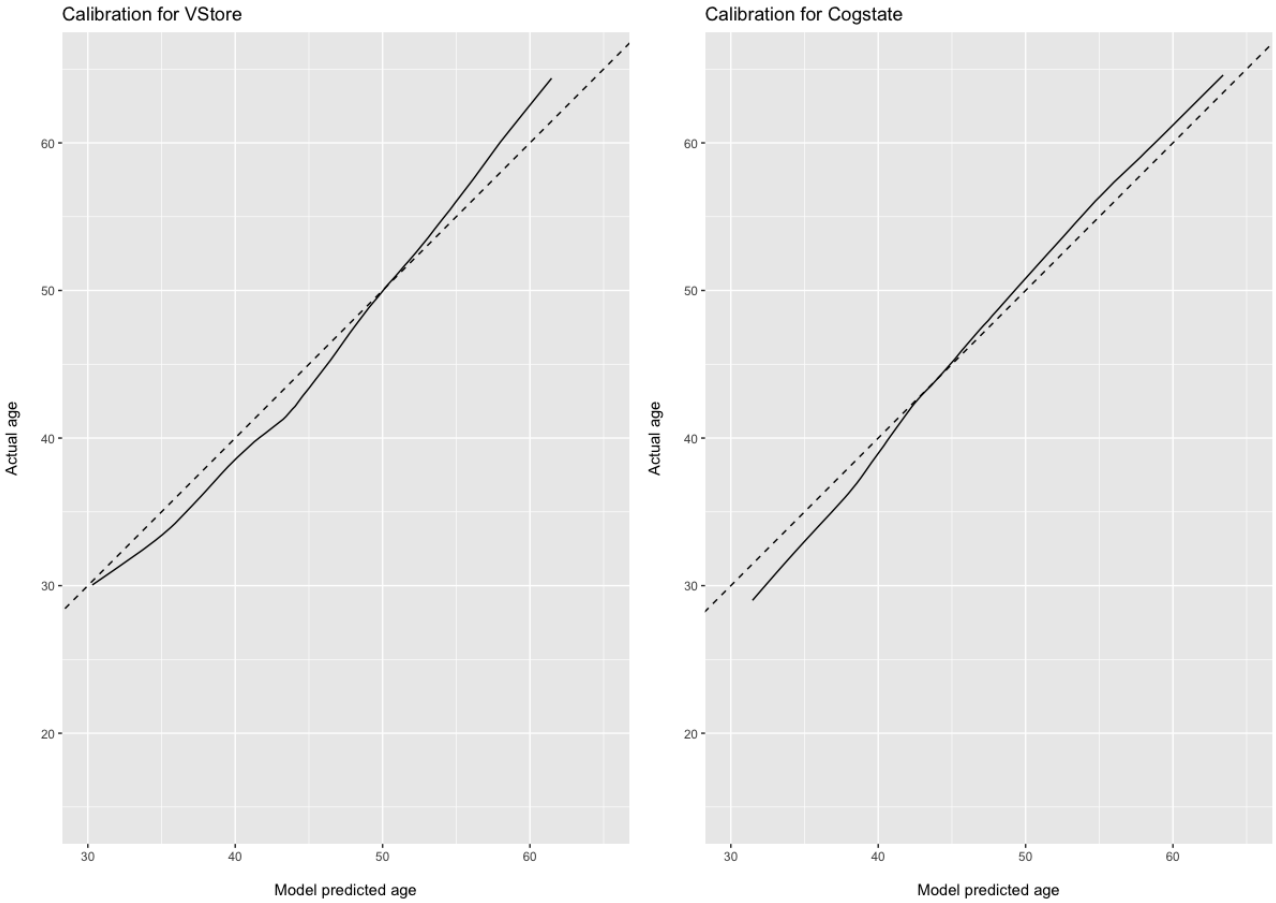


In the VStore model, coefficient values were as follows: VStore Recall=−0.586; VStore Find=7.882; VStore Select=5.284; VStore Pay=3.291; and VStore Coffee=4.526. In this model, the Find task is most strongly positively associated with increasing age, followed by the Select, Coffee, Pay, and Recall tasks.

In the Cogstate model, the coefficient values were as follows: DET=11.089; IDN=15.277; OCL=−2.563; ONB=12.038; TWO=−1.293; GMLT=0.032; CPAL=0.015; and ISLT=−0.245. In this model, the IDN task was most strongly positively associated with increasing age, followed by the ONB, DET, and TWO tasks.

For the DV age, we built 2 additional models using VStore and Cogstate outcomes as predictors with technological familiarity included as a covariate (Figure 3). With the TFQ added to the VStore model, the model fitting achieved an MSE of 162.9 (SE 17.50), selecting a total of 6 predictors and from cross-validating, this was attained from an optimal λ at 3.904 (Multimedia Appendix 17). With the TFQ added to the Cogstate model, we found an MSE of 175.4 (SE 22.12), selecting a total of 9 predictors obtained with an optimal λ at 2.904 (Multimedia Appendix 18).

Figure 3. VStore and Cogstate models predicting age with technological familiarity included. Dashed lines depict the age of the participants. The solid stroke shows the age predicted by the ridge regression models.



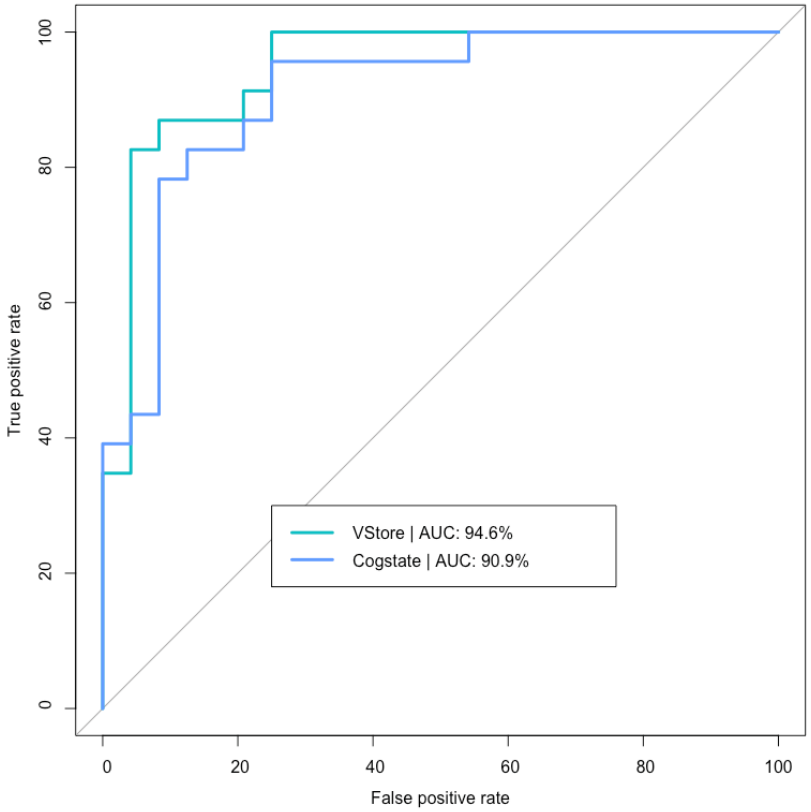
In the VStore model, coefficient values were as follows: VStore Recall=−0.617; VStore Find=7.298; VStore Select=5.422; VStore Pay=2.699; VStore Coffee=4.517, and TFQ=−0.279.

In the Cogstate model, coefficient values were as follows: DET=16.777; IDN=22.388; OCL=−3.544; ONB=15.451; TWO=2.034; GMLT=0.0438; CPAL=0.0243, ISLT=−0.479; and TFQ=−0.348.

Age Cohort Classification

Figure 4 shows the sensitivity of the VStore and Cogstate models in classifying age cohorts of 20-30 and 65-79 years of this study's sample population. VStore has a sensitivity of 87% and specificity of 91.7% at the optimal threshold of 0.55, whereas Cogstate has a sensitivity of 95.7% and specificity of 75% at the optimal threshold of 0.36. The difference between the 2 models was not statistically significant (Z=0.69, P=.49).

Figure 4. VStore and Cogstate models predicting age cohorts. AUC: area under the receiver operating characteristic curve.



Discussion

Principal Findings

The primary aim of this study was to establish which cognitive functions are engaged during a novel VR assessment, VStore. We found that a number of cognitive processes, as measured by Cogstate, contributed to the variance explained in VStore performance, suggesting that the VR task engages a range of key neuropsychological functions simultaneously. Indeed, the realistic nature of VStore precludes a simple one-to-one mapping between Cogstate domains and VStore outcomes. These findings provide preliminary information about VStore's construct validity and show that functional tasks embedded in VR may engage a greater range of cognitive domains than standard assessments because of their increased complexity and ability to resemble the demands of the real world [41].

As anticipated, VStore Recall was best explained by the Cogstate verbal learning task. VStore Find demonstrated a significant relationship with a number of predictors including attention, working memory, paired associate learning, age, and technological familiarity. VStore Select was explained by working memory, executive function, verbal learning, and technological familiarity. The more items participants could remember, the quicker they selected them on the self-checkout machine (verbal learning); attentional control (executive function) and temporary memorization of remaining items (working memory) were also required. VStore Pay engaged working memory, executive functions, and required processing speed. VStore Coffee was explained by visual and verbal learning, working memory, and age. Finally, the total time spent in VStore was best explained by Cogstate tasks measuring paired associate learning and working memory, in addition to the participants' age and technological familiarity, accounting for almost half of the variance in VStore performance.

The CPAL task of Cogstate is an episodic memory paradigm that involves visuospatial processing and indexes the ability to learn, store, and retrieve information. Paired associations may be especially important when finding items in a store, as this requires the retrieval of object representations from the shopping list, such as Cornflakes. Severe impairment in this domain has been linked to a number of neuropsychiatric conditions, including AD [42], and has been shown to be a valuable tool for the early detection of the disorder [43]. Deficits in paired associate learning have also been observed in schizophrenia and are linked to hippocampal volume loss [44].

Working memory, the temporary retention of information for manipulation and decision-making, is a key cognitive process in overall VStore performance. It is particularly relevant for the stages of the assessment where reviewing the shopping list is necessary to successfully carry out the next step of the task, such as finding an item or selecting it on the self-checkout machine. In support of the role of working memory in complex cognitive and functional assessments, factor analysis revealed that working memory was one of the latent variables of the VRFCAT, among problem solving and processing speed [45]. A decline in working memory has been reported in both AD and schizophrenia [46,47]. Working memory also declines as part of the normal aging process [48].

The ability of VStore to engage cognitive domains implicated in neuropsychiatric disorders and age-related cognitive decline

points to its potential in assessing functional cognition not only in healthy individuals but also in clinical populations. In this study, the total time spent in VStore increases with age; hence, age is a significant predictor of most VStore outcomes. However, this may partly be attributed to decrease in technological familiarity with age [49], which could also play a significant role in the outcome of digital assessments. Indeed, ridge regression revealed that the main VStore outcomes—Recall, Find, Select, Pay, and Coffee—provide a parsimonious model and can predict age accurately. Although we cannot make a direct comparison between VStore and Cogstate models, it is observed that Cogstate has a larger slope deviation from the identity line than VStore. Intriguingly, although the inclusion of technological familiarity made VStore model less precise, it did not alter the overall results. In contrast, the Cogstate model was markedly improved by the addition of technological familiarity. This may be because of the additional technological demands of the VStore setup, despite the intuitive nature of the task. The fact that the addition of technological familiarity did not improve the VStore model could be because the variance associated with technological skills was already captured, whereas for Cogstate, this was not the case. As technological familiarity decreased with age, we cannot rule out that VStore, similar to any other digital assessment, may potentially underestimate the cognitive abilities of older adults. As VR tools become more familiar, this relationship may reduce over time, and thus we recommend the assessment of technological familiarity in studies that include participants where these skills may vary.

Similar to these findings, receiver operating characteristic curve analysis revealed that VStore is highly accurate, sensitive, and specific to the classification of age cohorts, further supporting its potential use in the assessment of age-related cognitive decline. This is in line with previous research showing that age is a relevant factor in performance on VR assessments [49,50], potentially explicable by the decline in exploratory navigational abilities—a domain particularly vulnerable to the effects of aging [51]. Effective exploration and navigation are vital for completing VStore and are likely to engage relevant brain regions. Indeed, a key aim in designing the VStore Find task was to activate the place and grid cells in the hippocampus and entorhinal cortex [52]. Notably, this variable was the most strongly associated with increasing age, suggesting that spatial processing, as assessed by VStore, could be used to inform future normative data to detect below-average performance for specified age brackets with high sensitivity.

Limitations

There are several limitations to this study. First, the study sample had a high IQ on average, as expected from our highly educated cohort; hence, the sample may not be fully representative of the general population. This may be due to an oversampling from college students and a better-educated general population, and the use of the abbreviated IQ measure that relies on only 2 domains, verbal ability and matrix reasoning, and may generate inflated scores [53]. Nonetheless, we were able to include a range of IQ scores. Furthermore, relying on the AIC stepwise algorithm for model selection is not ideal, as it may be affected by several factors, such as the degree of correlation between predictors or the size of the sample, and thus may not be fully replicable [54]. Although theory-based model selection is preferable, given the novelty of the VR task, this was not possible on this occasion. In addition, although ridge regression models were cross-validated by optimizing λ, these models should be validated in an independent sample. Future research should also include measures of adverse VR effects; however, it is important to note here that no participant stopped the VR assessment because of cybersickness. Similarly, although there has been no functional capacity assessment developed for healthy adults, the inclusion of a proxy measure, such as the Cognitive Failures Questionnaire [55], would have been desirable. Finally, further research is required to confirm the construct validity of VStore and, most importantly, establish its test-retest reliability.

Conclusions

In conclusion, our findings suggest that VStore is a promising assessment that engages various cognitive functions, including those that tend to decline with age and during the development of neuropsychiatric disorders such as AD. Given that VStore simulates the complexity of everyday life in an ecologically valid environment, it may be suitable for evaluating functional cognition; however, further research is required to confirm this. VStore has theoretical advantages over other tests in being more engaging than traditional pen-and-paper and computerized batteries; it is fully immersive unlike other similar assessments, such as the VRFCAT, potentially increasing a psychological sensation of *being there* in a specific (virtual) surrounding [56], and thus enabling the assessment of real time cognitive and behavioral responses to that environment [57]. Furthermore, VStore provides complete experimental control, unlike the MET. Further research is urgently required to confirm age-related findings (ie, predictive validity in early cognitive decline) and establish its reliability and sensitivity to changes in cognition and functional capacity in both healthy and clinical samples.

Acknowledgments

The authors would like to thank Dr Dan W Joyce for his comments and suggestions on the data analysis. LAP is supported by the UK Medical Research Council (MR/N013700/1) and is a King's College London member of the Medical Research Council Doctoral Training Partnership in Biomedical Sciences. SSS is supported by the National Institute for Health Research Maudsley Biomedical Research Centre.

Authors' Contributions

LAP contributed to the study design and data collection and analysis and wrote the manuscript. MAM advised on data analysis, manuscript preparation, and review. JP contributed in data collection and read and edited the manuscript. CB contributed in data collection and read and edited the manuscript. JB contributed in VStore software development. TD advised on data analysis and read and edited the manuscript. EM cocontributed to the study design, data analysis, and manuscript preparation and review. SSS cocontributed to the study design, data analysis, and manuscript preparation and review.

Conflicts of Interest

SSS and EM created VStore, with contributions from LAP and technical development from Vitae VR Ltd. King's College London licensed its rights in VStore to Vitae VR Ltd. Authors LAP, EM, and SSS are entitled to a share of any revenue that King's College London may receive from commercialization of VStore by Vitae VR Ltd.

Multimedia Appendix 1
VStore courtyard.
[PNG File , 218 KB - jmir_v24i1e27641_app1.png]

Multimedia Appendix 2
VStore minimarket.
[PNG File , 232 KB - jmir_v24i1e27641_app2.png]

Multimedia Appendix 3
VStore shopping list.
[DOCX File , 13 KB - jmir_v24i1e27641_app3.docx]

Multimedia Appendix 4
VStore checkout.
[PNG File , 208 KB - jmir_v24i1e27641_app4.png]

Multimedia Appendix 5
Equipment specification.
[DOCX File , 13 KB - jmir_v24i1e27641_app5.docx]

Multimedia Appendix 6
VStore software.
[DOCX File , 12 KB - jmir_v24i1e27641_app6.docx]

Multimedia Appendix 7
VStore movement parameters.
[DOCX File , 12 KB - jmir_v24i1e27641_app7.docx]

Multimedia Appendix 8
Cogstate tasks.
[PDF File (Adobe PDF File), 447 KB - jmir_v24i1e27641_app8.pdf]

Multimedia Appendix 9
Technological Familiarity Questionnaire.
[DOCX File , 12 KB - jmir_v24i1e27641_app9.docx]

Multimedia Appendix 10
Mean and SD for VStore outcomes.
[DOCX File , 13 KB - jmir_v24i1e27641_app10.docx]

Multimedia Appendix 11
Mean and SD for Cogstate outcomes.
[DOCX File , 14 KB - jmir_v24i1e27641_app11.docx]

Multimedia Appendix 12
Bivariate correlations between Cogstate and VStore outcomes.
[DOCX File , 16 KB - jmir_v24i1e27641_app12.docx]

Multimedia Appendix 13
Linear regression model assumptions for VStore outcomes.
[DOCX File , 544 KB - jmir_v24i1e27641_app13.docx]

Multimedia Appendix 14
Bivariate correlations between the Technological Familiarity Questionnaire and Cogstate.
[DOCX File , 13 KB - jmir_v24i1e27641_app14.docx]

Multimedia Appendix 15
VStore ridge regression cross-validation fit without the Technological Familiarity Questionnaire.
[PNG File , 40 KB - jmir_v24i1e27641_app15.png]

Multimedia Appendix 16
Cogstate ridge regression cross-validation fit without the Technological Familiarity Questionnaire.
[PNG File , 43 KB - jmir_v24i1e27641_app16.png]

Multimedia Appendix 17
VStore ridge regression cross-validation fit with the Technological Familiarity Questionnaire.
[PNG File , 41 KB - jmir_v24i1e27641_app17.png]

Multimedia Appendix 18
Cogstate ridge regression cross-validation fit with the Technological Familiarity Questionnaire.
[PNG File , 41 KB - jmir_v24i1e27641_app18.png]

References

1. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):263-269 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]
2. Palmer BW, Heaton RK, Paulsen JS, Kuck J, Braff D, Harris MJ, et al. Is it possible to be schizophrenic yet neuropsychologically normal? Neuropsychology 1997 Jul;11(3):437-446. [doi: 10.1037//0894-4105.11.3.437] [Medline: 9223148]
3. Conradi HJ, Ormel J, de Jonge P. Presence of individual (residual) symptoms during depressive episodes and periods of remission: a 3-year prospective study. Psychol Med 2011 Jun;41(6):1165-1174. [doi: 10.1017/S0033291710001911] [Medline: 20932356]
4. Plassman BL, Langa KM, Fisher GG, Heeringa SG, Weir DR, Ofstedal MB, et al. Prevalence of cognitive impairment without dementia in the United States. Ann Intern Med 2008 Mar 18;148(6):427-434 [FREE Full text] [doi: 10.7326/0003-4819-148-6-200803180-00005] [Medline: 18347351]
5. Tatemichi TK, Desmond DW, Stern Y, Paik M, Sano M, Bagiella E. Cognitive impairment after stroke: frequency, patterns, and relationship to functional abilities. J Neurol Neurosurg Psychiatry 1994 Feb;57(2):202-207 [FREE Full text] [doi: 10.1136/jnnp.57.2.202] [Medline: 8126506]
6. Green MF. Cognitive impairment and functional outcome in schizophrenia and bipolar disorder. J Clin Psychiatry 2006 Oct 15;67(10):e12. [doi: 10.4088/jcp.1006e12]
7. McIntyre RS, Cha DS, Soczynska JK, Woldeyohannes HO, Gallaugher LA, Kudlow P, et al. Cognitive deficits and functional outcomes in major depressive disorder: determinants, substrates, and treatment interventions. Depress Anxiety 2013 Jun;30(6):515-527. [doi: 10.1002/da.22063] [Medline: 23468126]
8. Bosboom P, Alfonso H, Almeida O. Determining the predictors of change in quality of life self-ratings and carer-ratings for community-dwelling people with Alzheimer disease. Alzheimer Dis Assoc Disord 2013;27(4):363-371. [doi: 10.1097/WAD.0b013e318293b5f8] [Medline: 23632266]
9. Liu-Seifert H, Siemers E, Price K, Han B, Selzler KJ, Henley D, Alzheimer's Disease Neuroimaging Initiative. Cognitive impairment precedes and predicts functional impairment in mild Alzheimer's disease. J Alzheimer's Dis 2015;47(1):205-214 [FREE Full text] [doi: 10.3233/JAD-142508] [Medline: 26402769]

10. Fusar-Poli P, Deste G, Smieskova R, Barlati S, Yung AR, Howes O, et al. Cognitive functioning in prodromal psychosis: a meta-analysis. Arch Gen Psychiatry 2012 Jun;69(6):562-571. [doi: 10.1001/archgenpsychiatry.2011.1592] [Medline: 22664547]
11. Groves SJ, Douglas KM, Porter RJ. A systematic review of cognitive predictors of treatment outcome in major depression. Front Psychiatry 2018 Aug 28;9:382 [FREE Full text] [doi: 10.3389/fpsyt.2018.00382] [Medline: 30210368]
12. Keefe R, Buchanan R, Marder S, Schooler NR, Dugar A, Zivkov M, et al. Clinical trials of potential cognitive-enhancing drugs in schizophrenia: what have we learned so far? Schizophr Bull 2013 Mar;39(2):417-435 [FREE Full text] [doi: 10.1093/schbul/sbr153] [Medline: 22114098]
13. Salagre E, Solé B, Tomioka Y, Fernandes B, Hidalgo-Mazzei D, Garriga M, et al. Treatment of neurocognitive symptoms in unipolar depression: a systematic review and future perspectives. J Affect Disord 2017 Oct 15;221:205-221. [doi: 10.1016/j.jad.2017.06.034] [Medline: 28651185]
14. Sevigny J, Chiao P, Bussière T, Weinreb PH, Williams L, Maier M, et al. The antibody aducanumab reduces Aβ plaques in Alzheimer's disease. Nature 2016 Sep 01;537(7618):50-56. [doi: 10.1038/nature19323] [Medline: 27582220]
15. Wolf T, Edwards D, Giles G. Functional Cognition and Occupational Therapy: A Practical Approach to Treating Individuals With Cognitive Loss. Bethesda: AOTA Press; 2019.
16. Buchanan RW, Davis M, Goff D, Green MF, Keefe RS, Leon AC, et al. A summary of the FDA-NIMH-MATRICS workshop on clinical trial design for neurocognitive drugs for schizophrenia. Schizophr Bull 2005 Jan;31(1):5-19. [doi: 10.1093/schbul/sbi020] [Medline: 15888422]
17. Sabbagh MN, Hendrix S, Harrison JE. FDA position statement "early Alzheimer's disease: developing drugs for treatment, guidance for industry". Alzheimer’s Dement Transl Res Clin Interv 2019;5(1):13-19 [FREE Full text] [doi: 10.1016/j.trci.2018.11.004] [Medline: 31650002]
18. Bowie CR, Leung WW, Reichenberg A, McClure MM, Patterson TL, Heaton RK, et al. Predicting schizophrenia patients' real-world behavior with specific neuropsychological and functional capacity measures. Biol Psychiatry 2008 Mar 01;63(5):505-511 [FREE Full text] [doi: 10.1016/j.biopsych.2007.05.022] [Medline: 17662256]
19. Millan MJ, Agid Y, Brüne M, Bullmore ET, Carter CS, Clayton NS, et al. Cognitive dysfunction in psychiatric disorders: characteristics, causes and the quest for improved therapy. Nat Rev Drug Discov 2012 Feb 01;11(2):141-168. [doi: 10.1038/nrd3628] [Medline: 22293568]
20. Fervaha G, Zakzanis KK, Foussias G, Graff-Guerrero A, Agid O, Remington G. Motivational deficits and cognitive test performance in schizophrenia. JAMA Psychiatry 2014 Sep;71(9):1058-1065. [doi: 10.1001/jamapsychiatry.2014.1105] [Medline: 25075930]
21. Rosen WG, Mohs RC, Davis KL. A new rating scale for Alzheimer's disease. Am J Psychiatry 1984 Nov;141(11):1356-1364. [doi: 10.1176/ajp.141.11.1356] [Medline: 6496779]
22. Nuechterlein KH, Green MF, Kern RS, Baade LE, Barch DM, Cohen JD, et al. The MATRICS Consensus Cognitive Battery, part 1: test selection, reliability, and validity. Am J Psychiatry 2008 Feb;165(2):203-213. [doi: 10.1176/appi.ajp.2007.07010042] [Medline: 18172019]
23. Kern RS, Nuechterlein KH, Green MF, Baade LE, Fenton WS, Gold JM, et al. The MATRICS Consensus Cognitive Battery, part 2: co-norming and standardization. Am J Psychiatry 2008 Feb;165(2):214-220. [doi: 10.1176/appi.ajp.2007.07010043] [Medline: 18172018]
24. Morris JC. Clinical dementia rating: a reliable and valid diagnostic and staging measure for dementia of the Alzheimer type. Int. Psychogeriatr 2005 Jan 10;9(S1):173-176. [doi: 10.1017/s1041610297004870]
25. Patterson TL, Goldman S, McKibbin CL, Hughs T, Jeste DV. UCSD Performance-Based Skills Assessment: development of a new measure of everyday functioning for severely mentally ill adults. Schizophr Bull 2001;27(2):235-245. [doi: 10.1093/oxfordjournals.schbul.a006870] [Medline: 11354591]
26. Chaytor N, Schmitter-Edgecombe M. The ecological validity of neuropsychological tests: a review of the literature on everyday cognitive skills. Neuropsychol Rev 2003 Dec;13(4):181-197. [doi: 10.1023/b:nerv.0000009483.91468.fb] [Medline: 15000225]
27. Tarnanas I, Schlee W, Tsolaki M, Müri R, Mosimann U, Nef T. Ecological validity of virtual reality daily living activities screening for early dementia: longitudinal study. JMIR Serious Games 2013 Aug 06;1(1):e1 [FREE Full text] [doi: 10.2196/games.2778] [Medline: 25658491]
28. Shallice T, Burgess PW. Deficits in strategy application following frontal lobe damage in man. Brain 1991 Apr;114 (Pt 2):727-741. [doi: 10.1093/brain/114.2.727] [Medline: 2043945]
29. Keefe RS, Davis VG, Atkins AS, Vaughan A, Patterson T, Narasimhan M, et al. Validation of a computerized test of functional capacity. Schizophr Res 2016 Aug;175(1-3):90-96 [FREE Full text] [doi: 10.1016/j.schres.2016.03.038] [Medline: 27091656]
30. Ruse SA, Harvey PD, Davis VG, Atkins AS, Fox KH, Keefe RS. Virtual reality functional capacity assessment in schizophrenia: preliminary data regarding feasibility and correlations with cognitive and functional capacity performance. Schizophr Res Cogn 2014 Mar;1(1):e21-e26 [FREE Full text] [doi: 10.1016/j.scog.2014.01.004] [Medline: 25083416]
31. Negu A, Matu S, Sava FA, David D. Virtual reality measures in neuropsychological assessment: a meta-analytic review. Clin Neuropsychol 2016 Feb;30(2):165-184. [doi: 10.1080/13854046.2016.1144793] [Medline: 26923937]

32. Bird CM, Burgess N. The hippocampus and memory: insights from spatial processing. Nat Rev Neurosci 2008 Mar;9(3):182-194. [doi: 10.1038/nrn2335] [Medline: 18270514]
33. Moodley K, Minati L, Contarino V, Prioni S, Wood R, Cooper R, et al. Diagnostic differentiation of mild cognitive impairment due to Alzheimer's disease using a hippocampus-dependent test of spatial memory. Hippocampus 2015 Aug 26;25(8):939-951. [doi: 10.1002/hipo.22417] [Medline: 25605659]
34. Sheline YI, Wang PW, Gado MH, Csernansky JG, Vannier MW. Hippocampal atrophy in recurrent major depression. Proc Natl Acad Sci USA 1996 Apr 30;93(9):3908-3913 [FREE Full text] [doi: 10.1073/pnas.93.9.3908] [Medline: 8632988]
35. Heckers S. Neuroimaging studies of the hippocampus in schizophrenia. Hippocampus 2001;11(5):520-528. [doi: 10.1002/hipo.1068] [Medline: 11732705]
36. Vitae VR homepage. Vitae VR. URL: https://www.vitaevr.com/ [accessed 2021-12-11]
37. Zygouris S, Tsolaki M. Computerized cognitive testing for older adults: a review. Am J Alzheimers Dis Other Demen 2015 Feb 13;30(1):13-28 [FREE Full text] [doi: 10.1177/1533317514522852] [Medline: 24526761]
38. Wechsler Abbreviated Scale of Intelligence. The Psychological Corporation. 1999. URL: http://fcon_1000.projects.nitrc.org/indi/enhanced/assessments/wasi-ii.html [accessed 2021-12-11]
39. Venables W, Ripley B. Modern Applied Statistics with S. Cham: Springer; 2002.
40. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. J Stat Softw 2010;33(1):1-22 [FREE Full text] [Medline: 20808728]
41. Neguţ A, Matu S, Sava FA, David D. Task difficulty of virtual reality-based assessment tools compared to classical paper-and-pencil or computerized measures: a meta-analytic approach. Comput Human Behav 2016 Jan;54(C):414-424. [doi: 10.1016/j.chb.2015.08.029]
42. Egerházi A, Berecz R, Bartók E, Degrell I. Automated Neuropsychological Test Battery (CANTAB) in mild cognitive impairment and in Alzheimer's disease. Prog Neuropsychopharmacol Biol Psychiatry 2007 Apr 13;31(3):746-751. [doi: 10.1016/j.pnpbp.2007.01.011] [Medline: 17289240]
43. Fowler KS, Saling MM, Conway EL, Semple JM, Louis WJ. Paired associate performance in the early detection of DAT. J Int Neuropsychol Soc 2002 Jan 11;8(1):58-71. [doi: 10.1017/s1355617701020069]
44. Kéri S, Szamosi A, Benedek G, Kelemen O. How does the hippocampal formation mediate memory for stimuli processed by the magnocellular and parvocellular visual pathways? Evidence from the comparison of schizophrenia and amnestic mild cognitive impairment (aMCI). Neuropsychologia 2012 Dec;50(14):3193-3199. [doi: 10.1016/j.neuropsychologia.2012.10.010] [Medline: 23085125]
45. Ruse SA, Davis VG, Atkins AS, Krishnan KR, Fox KH, Harvey PD, et al. Development of a virtual reality assessment of everyday living skills. J Vis Exp 2014 Apr 23(86) [FREE Full text] [doi: 10.3791/51405] [Medline: 24798174]
46. Baddeley AD, Bressi S, Della Sala S, Logie R, Spinnler H. The decline of working memory in Alzheimer's disease. A longitudinal study. Brain 1991 Dec;114 (Pt 6):2521-2542. [doi: 10.1093/brain/114.6.2521] [Medline: 1782529]
47. Forbes NF, Carrick LA, McIntosh AM, Lawrie SM. Working memory in schizophrenia: a meta-analysis. Psychol Med 2009 Jun;39(6):889-905. [doi: 10.1017/S0033291708004558] [Medline: 18945379]
48. Chai WJ, Abd Hamid AI, Abdullah JM. Working memory from the psychological and neurosciences perspectives: a review. Front Psychol 2018 Mar 27;9:401 [FREE Full text] [doi: 10.3389/fpsyg.2018.00401] [Medline: 29636715]
49. Olson KE, O'Brien MA, Rogers WA, Charness N. Diffusion of technology: frequency of use for younger and older adults. Ageing Int 2011 Mar 28;36(1):123-145 [FREE Full text] [doi: 10.1007/s12126-010-9077-9] [Medline: 22685360]
50. Oliveira CR, Lopes Filho BJ, Esteves CS, Rossi T, Nunes DS, Lima MM, et al. Neuropsychological assessment of older adults with virtual reality: association of age, schooling, and general cognitive status. Front Psychol 2018;9:1085 [FREE Full text] [doi: 10.3389/fpsyg.2018.01085] [Medline: 30008689]
51. Yamamoto N, Degirolamo GJ. Differential effects of aging on spatial learning through exploratory navigation and map reading. Front Aging Neurosci 2012 Jun 12;4:14 [FREE Full text] [doi: 10.3389/fnagi.2012.00014] [Medline: 22701423]
52. Hafting T, Fyhn M, Molden S, Moser M, Moser EI. Microstructure of a spatial map in the entorhinal cortex. Nature 2005 Aug 11;436(7052):801-806. [doi: 10.1038/nature03721] [Medline: 15965463]
53. Axelrod BN. Validity of the Wechsler abbreviated scale of intelligence and other very short forms of estimating intellectual functioning. Assessment 2002 Mar 26;9(1):17-23. [doi: 10.1177/1073191102009001003] [Medline: 11911230]
54. Derksen S, Keselman H. Backward, forward and stepwise automated subset selection algorithms: frequency of obtaining authentic and noise variables. Br J Math Stat Psychol 1992;45(2):265-282. [doi: 10.1111/j.2044-8317.1992.tb00992.x]
55. Broadbent D, Cooper P, FitzGerald P, Parkes K. The Cognitive Failures Questionnaire (CFQ) and its correlates. Br J Clin Psychol 1982 Feb;21(1):1-16. [doi: 10.1111/j.2044-8260.1982.tb01421.x] [Medline: 7126941]
56. Slater M. Presence and emotions. Cyberpsychol Behav 2004 Feb;7(1):121; author reply 123. [doi: 10.1089/109493104322820200] [Medline: 15006177]
57. Eichenberg C, Wolters C. Virtual realities in the treatment of mental disorders: a review of the current state of research. Virtual Reality Psychol Med Pedagogical App 2012. [doi: 10.5772/50094]

Abbreviations

AD: Alzheimer disease
AIC: Akaike Information Criterion
CPAL: Continuous Paired Associate Learning
DET: Detection
DV: dependent variable
GMLT: Groton Maze Learning Task
IDN: Identification
ISLT: International Shopping List Task
IV: independent variable
MET: Multiple Errands Test
MSE: mean squared error
OCL: One Card Learning
ONB: One-Back
TFQ: Technological Familiarity Questionnaire
TWO: Two-Back
VR: virtual reality
VRFCAT: Virtual Reality Functional Capacity Assessment Tool

Edited by R Kukafka, G Eysenbach; submitted 01.02.21; peer-reviewed by H Huygelier, R Lundin; comments to author 10.05.21; revised version received 05.07.21; accepted 05.10.21; published 26.01.22.

Please cite as:
Porffy LA, Mehta MA, Patchitt J, Boussebaa C, Brett J, D'Oliveira T, Mouchlianitis E, Shergill SS
A Novel Virtual Reality Assessment of Functional Cognition: Validation Study
J Med Internet Res 2022;24(1):e27641
URL: https://www.jmir.org/2022/1/e27641
doi:10.2196/27641
PMID:35080501

Original Paper

Intensive Care Unit–Specific Virtual Reality for Critically Ill Patients With COVID-19: Multicenter Randomized Controlled Trial

Johan H Vlake[1,2], BSc; Jasper van Bommel[1], MD, PhD; Evert-Jan Wils[2], MD, PhD; Joe Bienvenu[3], MD, PhD; Merel E Hellemons[4], MD, PhD; Tim IM Korevaar[5], MD, PhD; Anna FC Schut[6], MD, PhD; Joost AM Labout[7], MD, PhD; Lois LH Schreuder[1], MD; Marten P van Bavel[1], BSc; Diederik Gommers[1], MD, PhD; Michel E van Genderen[1], MD, PhD

[1]Department of Intensive Care, Erasmus MC, Rotterdam, Netherlands

[2]Department of Intensive Care, Franciscus Gasthuis & Vlietland, Rotterdam, Netherlands

[3]Department of Psychiatry and Behavioral Sciences, John Hopkins University School of Medicine, Baltimore, MD, United States

[4]Department of Pulmonology, Erasmus MC, Rotterdam, Netherlands

[5]Department of Internal Medicine, Academic Centre for Thyroid Diseases, Erasmus MC, Rotterdam, Netherlands

[6]Department of Intensive Care, Ikazia hospital, Rotterdam, Netherlands

[7]Department of Intensive Care, Maasstad hospital, Rotterdam, Netherlands

Corresponding Author:
Michel E van Genderen, MD, PhD
Department of Intensive Care, Erasmus MC
Dr Molewaterplein 40
Rotterdam, 3015 GD
Netherlands
Phone: 31 107039881
Email: m.vangenderen@erasmusmc.nl

Abstract

Background: Although psychological sequelae after intensive care unit (ICU) treatment are considered quite intrusive, robustly effective interventions to treat or prevent these long-term sequelae are lacking. Recently, it was demonstrated that ICU-specific virtual reality (ICU-VR) is a feasible and acceptable intervention with potential mental health benefits. However, its effect on mental health and ICU aftercare in COVID-19 ICU survivors is unknown.

Objective: This study aimed to explore the effects of ICU-VR on mental health and on patients' perceived quality of, satisfaction with, and rating of ICU aftercare among COVID-19 ICU survivors.

Methods: This was a multicenter randomized controlled trial. Patients were randomized to either the ICU-VR (intervention) or the control group. All patients were invited to an COVID-19 post-ICU follow-up clinic 3 months after hospital discharge, during which patients in the intervention group received ICU-VR. One month and 3 months later (4 and 6 months after hospital discharge), mental health, quality of life, perceived quality, satisfaction with, and rating of ICU aftercare were scored using questionnaires.

Results: Eighty-nine patients (median age 58 years; 63 males, 70%) were included. The prevalence and severity of psychological distress were limited throughout follow-up, and no differences in psychological distress or quality of life were observed between the groups. ICU-VR improved satisfaction with (mean score 8.7, SD 1.6 vs 7.6, SD 1.6 [ICU-VR vs control]; t_{64}=–2.82, P=.006) and overall rating of ICU aftercare (mean overall rating of aftercare 8.9, SD 0.9 vs 7.8, SD 1.7 [ICU-VR vs control]; t_{64}=–3.25; P=.002) compared to controls. ICU-VR added to the quality of ICU aftercare according to 81% of the patients, and all patients would recommend ICU-VR to other ICU survivors.

Conclusions: ICU-VR is a feasible and acceptable innovative method to improve satisfaction with and rating of ICU aftercare and adds to its perceived quality. We observed a low prevalence of psychological distress after ICU treatment for COVID-19, and ICU-VR did not improve psychological recovery or quality of life. Future research is needed to confirm our results in other critical illness survivors to potentially facilitate ICU-VR's widespread availability and application during follow-up.

Trial Registration: Netherlands Trial Register NL8835; https://www.trialregister.nl/trial/8835

International Registered Report Identifier (IRRID): RR2-10.1186/s13063-021-05271-z

(J Med Internet Res 2022;24(1):e32368) doi:10.2196/32368

KEYWORDS

SARS-CoV-2; intensive care; post-intensive care syndrome; virtual reality; quality of life; satisfaction; COVID-19

Introduction

The increase in the survival of critically ill patients admitted to the intensive care unit (ICU) in the last few decades has revealed the effect of ICU treatment on quality of life [1-3]. Up to one-third of "general" ICU survivors experience a poor quality of life, predominantly owing to psychological sequelae such as anxiety, depression, and posttraumatic stress disorder (PTSD) [4-7]. These psychological impairments comprise the psychological component of the postintensive care syndrome (PICS); they are common and can last months to even years after patient ICU discharge [4,8,9]. Consequently, there is a need for post-ICU care.

As the demand for ICU beds and critical care services has skyrocketed during the current COVID-19 outbreak, so would that of ICU aftercare. As such, health care services will have to adapt rapidly to an anticipated surge of post-ICU care, and this will place an enormous strain on acute services [10]. The current pandemic is highlighting the urgency for a multimodal follow-up program and the need for patient-focused innovative solutions [11]. Difficulty accessing in-person clinics is a key barrier in the development of an ICU follow-up program and is probably hindered more owing to current COVID-19 regulations [12]. While post-ICU care has been recognized as a fundamental part of ICU care by the critical care community, effective interventions and guidelines are lacking, and evidence for the effectiveness of post-ICU programs has not been established [4,13,14]. The unmet need for information and the increasing importance of satisfaction as an important quality indicator are two major denominators that could explain the aforementioned and should, as such, be taken into account in developing post-ICU care in the current era [15,16].

Recovery from COVID-19 could have the same multifaceted problems that occur after sepsis and other critical illnesses [17]. Recently, virtual reality (VR) was demonstrated to be a useful technique to improve post-ICU mental health in sepsis survivors and could also be safely used and implemented in post-ICU COVID-19 care [18-20]. As such, we hypothesized that an ICU-specific virtual reality (ICU-VR) intervention could improve the satisfaction with and rating of ICU aftercare and could contribute to psychological recovery. The aims of this study were therefore to explore the effects of ICU-VR on mental health and on patients' perceived quality and satisfaction with and rating of ICU aftercare among COVID-19 ICU survivors.

Methods

Study Design

This multicenter, open-label, randomized controlled trial was conducted in a university teaching hospital and in 3 university-affiliated secondary care hospitals. Patients were included from June 2020 to February 2021 and were followed-up for 6 consecutive months. The study protocol was approved by the Medical Ethics Committee of the Erasmus Medical Centre, Rotterdam, and the participating centers' institutional review boards (NL73667.078.20, approved June 10, 2020) and has previously been published [21].

Participants

All consecutive adult (≥18 years) patients who were treated in an ICU of one of the participating hospitals and visited the COVID-19 post-ICU follow-up clinic were eligible for inclusion. COVID-19 was diagnosed on the basis of a positive finding on reverse transcription–polymerase chain reaction (RT–PCR) for SARS-CoV-2. Exclusion criteria were primary neurological impairments or documented active psychiatric diseases, an inability to understand the Dutch language, absence of a formal home address, and participation in other interventional trials that could confound the primary outcome. Patients and the public were not involved in the design, conduct, reporting, or dissemination plans of this study. A former ICU patient was involved in the development of the ICU-VR intervention.

Randomization and Masking

Patients were randomly assigned to either the ICU-VR (intervention) group or the control group at a 1:1 ratio, using a centralized internet-based randomization procedure by the study site's principal investigator or a representative (Castor Electronic Data Capture [EDC]). Patients were randomized in a simple manner without stratification. The investigators were unaware of the assignment sequence. Owing to the nature of the intervention, blinding of patients and investigators was not possible.

Intervention

All patients were invited to a COVID-19 post-ICU follow-up clinic as part of regional standard care. During this visit, patients had a 60-minute-long consultation with an intensivist and an ICU nurse, during which the ICU treatment was reviewed, and patients were screened for PICS-related impairments and referred to an appropriate health care worker, if appropriate.

Patients in the ICU-VR group received the ICU-VR intervention once during this visit. ICU-VR is explained in depth elsewhere. In short, it was developed by an interdisciplinary team that included intensivists, ICU nurses, a psychologist, a psychiatrist, an investigator, and former ICU patients and was previously demonstrated to be safe and feasible [18,22]. ICU-VR consists of a 14-minute-long informational video that can be watched using VR, in which the patient is exposed to the ICU environment and receives voice-over explanations regarding different facets of the surrounding ICU and ICU treatment. ICU-VR consists of 6 scenes: (1) The ICU physician and nurse welcome the patient in front of the ICU. After being brought to and installed in the ICU, explanations are given (2) about the surveillance monitor, medication pumps, intubation (including tracheal tube suction), mechanical ventilation, and prone

positioning; (3) about intravenous drips and lines and tracheotomy, including its procedures; (4) about the treatment team taking care of the patient; (5) about isolation measures and personal protection equipment; and (6) about COVID-19 [19,21]. The script and the YouTube version can be found elsewhere [21,23]. The ICU-VR intervention was watched using head-mounted display–VR glasses (Oculus Go) in combination with headphones.

Study Procedures

All COVID-19 ICU survivors were invited to the hospital's COVID-19 post-ICU follow-up clinic 3 months after hospital discharge as part of regional standard care. One month prior to this visit, eligible patients were sent a study information brochure, and 2 weeks later, patients were contacted by telephone by a member of the study team to explain the study procedures. During their follow-up clinic visit, consent was obtained, and patients were randomized.

Patients randomized to the ICU-VR group received the ICU-VR intervention once during the concordant follow-up clinic visit, whereas patients randomized to the control group did not receive ICU-VR. Aside from the ICU-VR intervention, there were no differences between the study groups. These results are part of a larger study evaluating the long-term effects of ICU-VR after 6 months and the effect of VR crossover [21].

Prior to the follow-up clinic visit and 4 and 6 months after hospital discharge—that is, 1 and 3 months after the COVID-19 post-ICU follow-up clinic visit—psychological distress and quality of life were assessed. Six months after hospital discharge, all patients were asked about their satisfaction with and rating of ICU care and aftercare, and patients in the intervention group were asked about their perspectives on ICU-VR.

Outcomes

Primary outcomes were PICS-related psychological distress and quality of life up to 6 months after hospital discharge and were mandatory parts of the questionnaire.

Psychological distress was expressed as the prevalence and severity of PTSD, anxiety, and depression-related symptoms assessed using the Impact of Event Scale-Revised (IES-R; PTSD) and the Hospital Anxiety and Depression Scale (HADS; anxiety and depression), respectively [24,25]. The IES-R is a self-reported measurement consisting of 22 items that assesses subjective distress caused by a traumatic event and has previously been validated in ICU survivors [24,26,27]. It provides a total score ranging from 0 to 88, with higher scores indicating more severe symptoms. It also provides subscale scores to assess symptoms of intrusion, avoidance, and hyperarousal, which is the sum of all items in each section. An IES-R total score of ≥34 is considered the optimal cutoff for PTSD [28]. The HADS consists of 14 items and is commonly used to determine the levels of anxiety and depression that a patient is experiencing and has been validated in critical illness survivors [29-31]. Seven of the items relate to anxiety, 7 relate to depression, and each question is answered on a 4-point Likert scale. A sum score of ≥8 (ranging from 0 to 21, with higher scores indicating more severe symptoms) on either the depression or the anxiety subscale, is classified as clinically meaningful depression and anxiety, respectively [29].

Quality of life was assessed using the Short-Form 36 (SF-36) and the European Quality of Life, 5 Dimensions (EQ-5D) questionnaires [32,33]. The EQ-5D and SF-36 have been validated and tested in the ICU and have been recommended for use in critical care medicine [34-38]. The EQ-5D measures quality of life in 5 dimensions (mobility, self-care, usual activities, pain or discomfort, and anxiety or depression). In each domain, patients are asked if they experience no, slight, moderate, severe, or extreme problems, from which the weight of a health state can be computed, ranging from –0.446 (worst quality of life) to 1.000 (best quality of life) [39]. Additionally, patients score their current subjective health on a visual analog scale, ranging from 0 (worst health imaginable) to 100 (best health imaginable). SF-36 is a 36-item, patient-reported survey of health and health-related quality of life (HRQoL). It consists of 8 scaled scores, which are the weighted sums of the questions in their section, and a scale for health change. Each scale is directly transformed to a scale ranging from 0 (worst score) to 100 (best score) on the assumption that each question carries an equal weight. The 8 sections are physical functioning, social functioning, physical role functioning, emotional role functioning, emotional well-being, vitality, bodily pain, and general health perception [40]. In addition to these scales, mental and physical component scores can be calculated, which represent a patient's mental and physical health state. These scores are computed so that the mean is 50 (SD 10) for the general population [41,42].

Patients' perceived quality of and patients' satisfaction with and rating of ICU aftercare were assessed using a novel questionnaire, and the questions were nonmandatory to answer. The questionnaire was based on the Patient Satisfaction Questionnaire and Family Satisfaction with ICU Care tools, altered to the needs of this study [43-45]. This questionnaire consisted of 21 items and was categorized into five sections: perspectives on the added value of ICU-VR to ICU care and ICU aftercare (8 questions), perspectives on the timing and number of sessions (3 questions), overall perspective on the ICU-VR intervention (3 questions), perspectives on the content of the ICU-VR intervention (3 questions), and perspectives on the effect of the ICU-VR intervention (4 questions; Multimedia Appendix 1). The first section was (partly) answered by all patients, irrespective of the randomization allocation, and the other sections were answered only by patients randomized to the ICU-VR group. All questions could be answered on a 10-point Likert scale, ranging from 1 (not at all) to 10 (very much), except for perspectives on the timing and number of sessions. The questionnaire was administered by telephone.

Baseline characteristics and survival were determined through patient record analysis. Additional demographics, such as educational level and preadmission employment status, were assessed using follow-up questionnaires.

Statistical Analysis

Based on a previous pilot study examining the feasibility, safety, and clinical relevance of sepsis ICU-VR (Cohen *d* effect size=0.77), we assumed the effect estimates of ICU-VR to be

similar in this study [20]. Using a 2-sided α value of .05, a power of .80, a 1:1 randomization, and an expected loss to follow-up of 20%, we aimed to include a minimum of 80 patients, with 40 patients in each study group.

Baseline demographics and treatment-related characteristics were quantified using descriptive statistics. Continuous variables are expressed as median (IQR) or mean (SD) values, depending on their distribution. Categorical variables are presented as absolute numbers and relative frequencies.

Differences between study groups in continuous variables, such as the IES-R sum score, the HADS anxiety and depression scores, the SF-36 subscales and the EQ-5D utility score, at several follow-up time points were analyzed using a mixed-effects linear regression model with a random intercept for each study site. Differences in continuous outcomes at the 3-month follow-up time point were adjusted by adding the 3-month outcome as an independent variable in the mixed-effects linear regression model. Patients were categorized on the basis of clinically meaningful cutoffs for the IES-R sum score and the HADS anxiety and depression scores. Differences in categorical variables between study groups at several follow-up time points were analyzed using a mixed-effects logistic regression model with a random effect for each site. Differences in categorical outcomes at the 3-month follow-up time point were adjusted by adding the 3-month outcome as an independent variable in the mixed-effects logistic regression model. Differences in continuous or categorical variables throughout follow-up were analyzed using a mixed-effects linear or logistic regression model with time, randomization, and a random intercept or slope for each individual and each study site as appropriate. Differences in linear or categorical outcomes at the 3-month follow-up time point were added to the mixed-effects linear or logistic model as independent variables to adjust for that difference.

Outcomes of the mixed-effects linear regression models are reported as coefficient (95% CI) values, which implies the estimated mean difference, and outcomes of the mixed-effects logistic regression models are reported as odds ratios (ORs) with corresponding 95% CI values.

All data were gathered using Castor EDC. All analyses were performed using SPSS (version 24.0; SPSS Inc) and R for Statistics (R Foundation for Statistical Computing). A *P* value of ≤.05 was considered statistically significant.

Data Sharing

All data sets created during this study are available upon reasonable request by the corresponding author.

Results

Results Overview

A total of 147 patients visited the COVID-19 post-ICU follow-up clinic, of whom 89 were enrolled (inclusion rate: 61%): 45 patients in the ICU-VR group and 44 patients in the control group (Figure 1). All patients in the ICU-VR group completed the ICU-VR intervention, and no adverse events were reported. Baseline demographics and treatment-related characteristics were well balanced between groups (Table 1). The mean age was 58 (SD 11) years, 63 patients (71%) were male, and the median ICU length of stay was 17 (IQR 9-29) days. Table 1 shows baseline demographics and treatment-related characteristics.

Figure 1. CONSORT (Consolidated Standards of Reporting Trials) flow diagram of the study. ICU-VR: intensive care unit–specific virtual reality.

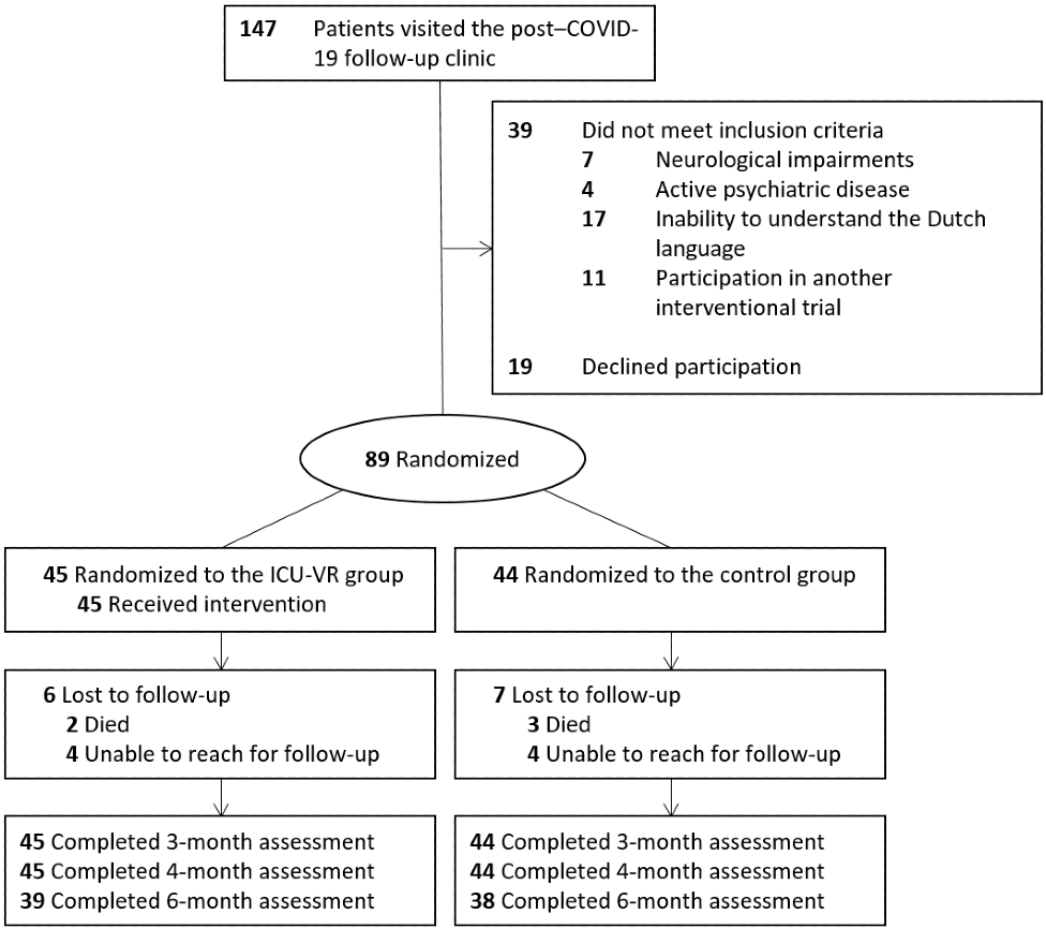


Table 1. Baseline demographics and treatment-related characteristics.

Characteristics[a]	ICU-VR[b] group (intervention) (n=45)	Control group (n=44)
Baseline demographics		
Age (years), median (IQR)	61 (54-65)	59 (51-65)
Males, n (%)	35 (78)	28 (36)
BMI, median (IQR)	27.6 (25.3-31.1)	28.0 (25.3-31.2)
Participants with a BMI of >30, n (%)	14 (31)	18 (41)
Educational level, n (%)		
Primary education	14 (31)	13 (30)
Intermediate vocational education	15 (33)	20 (46)
Higher vocational education	13 (29)	7 (16)
Academic education	3 (7)	4 (9)
Employment status, employed, n (%)	23 (51)	21 (48)
Treatment-related characteristics		
Length of stay in the intensive care unit (days), median (IQR)	14 (9-25)	14 (7-28)
Length of hospital stay (days), median (IQR)	22 (12-32)	24 (13-40)
Mechanical ventilation, n (%)	41 (91)	38 (86)
Duration (hours), median (IQR)	227 (169-343)	383 (206-465)
Highest positive end-respiratory pressure (cm H_2O), median (IQR)	21 (17-28)	20 (16-25)
Lowest fraction of inspired oxygen (%), median (IQR)	28 (24-30)	25 (22-30)
Lowest ratio of arterial oxygen (mm Hg), median (IQR)	0.11 (0.09-0.23)	0.11 (0.09-0.18)
Prone positioning, n (%)	35 (77)	36 (82)
Medication		
Received noradrenaline, n (%)	37 (82)	35 (80)
Noradrenaline dose (µg/kg/minute), median (IQR)	0.17 (0.10-0.30)	0.14 (0.08-0.29)
Duration of noradrenaline use (hours), median (IQR)	186 (32-249)	167 (96-349)
Received midazolam, n (%)	35 (78)	33 (75)
Midazolam dose (mg/kg/hour), median (IQR)	0.59 (0.43-0.71)	0.51 (0.39-0.66)
Duration of midazolam use (hours), median (IQR)	20 (13-93)	20 (13-36)
Received remifentanil, n (%)	32 (71)	35 (80)
Remifentanil dose (µg/kg/hour), median (IQR)	14 (10-16)	14 (6-18)
Duration of remifentanil use (hours), median (IQR)	33 (22-80)	32 (23-72)
Received sufentanil, n (%)	26 (58)	28 (63)
Sufentanil dose (µg/kg/hour), median (IQR)	0.55 (0.34-0.83)	0.60 (0.38-0.70)
Duration of sufentanil use (hours), median (IQR)	8 (1-13)	10 (6-14)
Received rocuronium, n (%)	22 (49)	16 (36)
Rocuronium dose (mg/kg/hour), median (IQR)	0.39 (0.05-0.77)	0.32 (0.01-0.60)
Duration of rocuronium use (hours), median (IQR)	22 (0-28)	17 (0-22)
Illness severity scores		
Simplified Acute Physiology Score (version 2), median (IQR)	31 (26-36)	31 (26-35)
Acute Physiology and Chronic Health Evaluation (version 4) score, median (IQR)	49 (38-60)	49 (42-59)
Admission Sequential Organ Failure Assessment score, median (IQR)	2 (1-6)	2 (1-4)

Characteristics[a]	ICU-VR[b] group (intervention) (n=45)	Control group (n=44)
Highest Sequential Organ Failure Assessment score, median (IQR)	8 (6-10)	7 (6-9)

[a]Baseline demographics and treatment-related characteristics were obtained at 3 months after hospital discharge via digital patient records.

[b]ICU-VR: intensive care unit-specific virtual reality.

Psychological Component of the PICS

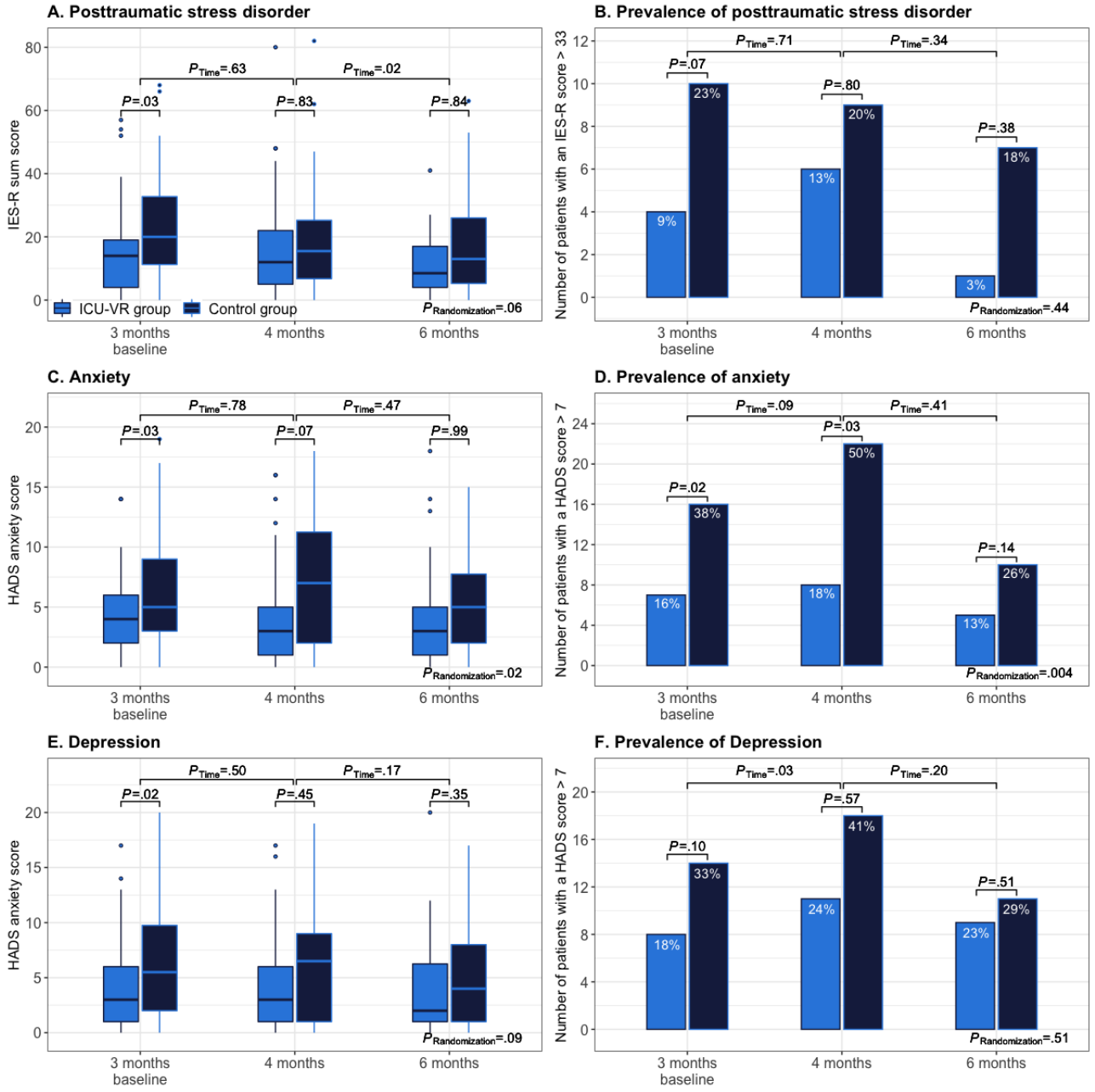
At the 3-month follow-up time point, a total of 31 of 89 patients (34%) reported psychologic distress, with 10 patients (22%) in the ICU-VR group and 21 patients (47%) in the control group (OR 3.5, 95% CI 1.4-8.9, *P*<.01). At 4 months, 38 patients (43%) reported psychological distress, with 12 patients (27%) in the ICU-VR group and 26 patients (59%) in the control group (OR 3.0, 95% CI 0.8-11.9, *P*=.11). At 6 months, 24 patients (31%) reported psychological distress, with 9 patients (23%) in the ICU-VR group and 15 patients (39%) in the control group (OR 0.7, 95% CI 0.2-2.9, *P*=.60).

At the 3-month follow-up time point, 4 patients (9%) in the ICU-VR group and 10 patients (22%) in the control group reported probable PTSD (OR 3.2, 95% CI 0.9-11.1, *P*=.07; Figure 2B). During follow-up, no differences were observed in PTSD scores or the proportion of patients who reported probable PTSD between randomization allocations (Figures 2A and 2B). Throughout follow-up, the PTSD score remained similar at 4 months (β=–.60, 95% CI –3.2 to 1.9, *P*=.63) after hospital discharge but improved at 6 months after hospital discharge (β=–3.1, 95% CI –5.8 to –0.4, *P*=.02). However, this improvement was independent of the randomization group (β=5.4, 95% CI –0.2 to 11.1, *P*=.06).

At the 3-month follow-up time point, 7 patients (16%) in the ICU-VR group and 16 patients (38%) in the control group reported probable anxiety (OR 3.3, 95% CI 1.2-9.3, *P*=.02; Figure 2D). Four months after hospital discharge, ICU-VR resulted in fewer patients with probable anxiety (n=8, 18% vs 22, 50%; OR 3.8, 95% CI 1.1-12.7; *P*=.03; Figure 2C) but not lower anxiety scores (median HADS anxiety score 3, IQR 1-5 vs 7, IQR 2-11; β=1.4, 95% CI –0.1 to 3.0; *P*=.07). There were no differences at 4 or 6 months after hospital discharge (Figures 2C and 2D). No natural decline in anxiety was observed, and the severity of anxiety and the prevalence of probable anxiety were lower in the ICU-VR group throughout the follow-up.

At the 3-month follow-up time point, 8 (18%) patients in the ICU-VR group and 14 (33%) patients in the control group reported probable depression (OR 2.3, 95% CI 0.9-6.3, *P*=.10; Figure 2F). Throughout the follow-up, no difference in the depression scores or the proportion of patients reporting probable depression was observed (Figures 2E and 2F). The severity of depression remained similar throughout the follow-up period.

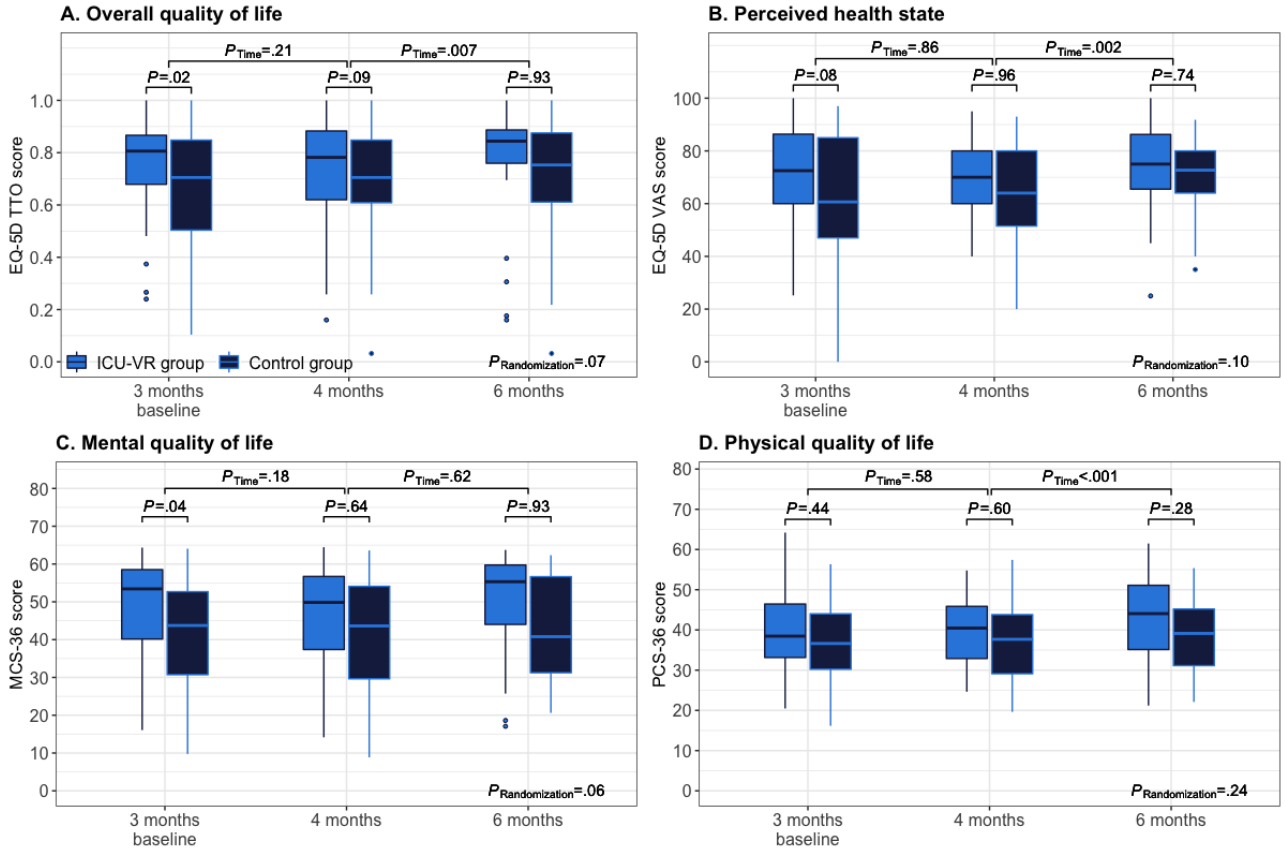
Figure 2. Psychological outcomes. Boxplots of the severity of posttraumatic stress disorder (A), anxiety (C), and depression (E) and bar plots of the prevalence of posttraumatic stress disorder (B), anxiety (D), and depression (F). Posttraumatic stress disorder was assessed using the IES-R, and a sum score of ≥33 was considered as posttraumatic stress disorder being prevalent; anxiety and depression were assessed using the HADS, and a score of ≥8 on either the anxiety or depression scale was considered anxiety and depression being prevalent, respectively. Differences between randomization groups at each follow-up time point and between follow-up time points (p, Time) and throughout the follow-up (p, Randomization) were analyzed using mixed-effects linear (severity) or logistic (prevalence) regression models. HADS: Hospital Anxiety and Depression Scale, IES-R: Impact of Event Scale-Revised.



Health-Related Quality of Life

The overall health-related quality of life, mental health–related quality of life, and physical health–related quality of life are depicted in Figure 3**.** Throughout the follow-up period, the overall, mental, and physical HRQoL remained similar until 4 months but improved at 6 months after hospital discharge, while overall quality of life, outcomes of individual EQ-5D domains, and subscales of the SF-36 score differed between groups during the follow-up period (Figures 3A-D, Multimedia Appendix 2 and Multimedia Appendix 3).

Figure 3. Quality of life outcomes. Boxplots of the overall quality of life (A), perceived health state (B), mental quality of life (C), and physical quality of life (D). Overall quality of life was expressed as the EQ-5D TTO score, the perceived health state as the EQ-5D VAS score, and the mental and physical quality of life as the mental and physical component scales of the SF-36, respectively. Differences between randomization groups at each follow-up time point and between follow-up time points (p, Time) and throughout the follow-up (p, Randomization) were analyzed using mixed-effects linear (severity) or logistic (prevalence) regression models. EQ-5D: European Quality of Life, 5 dimensions, ICU-VR: intensive care unit–virtual reality, MCS-36: Mental Component Summary, 36 items, PCS-36: Physical Component Summary, 36 items, TTO: trade time-off, VAS: visual analog scale.

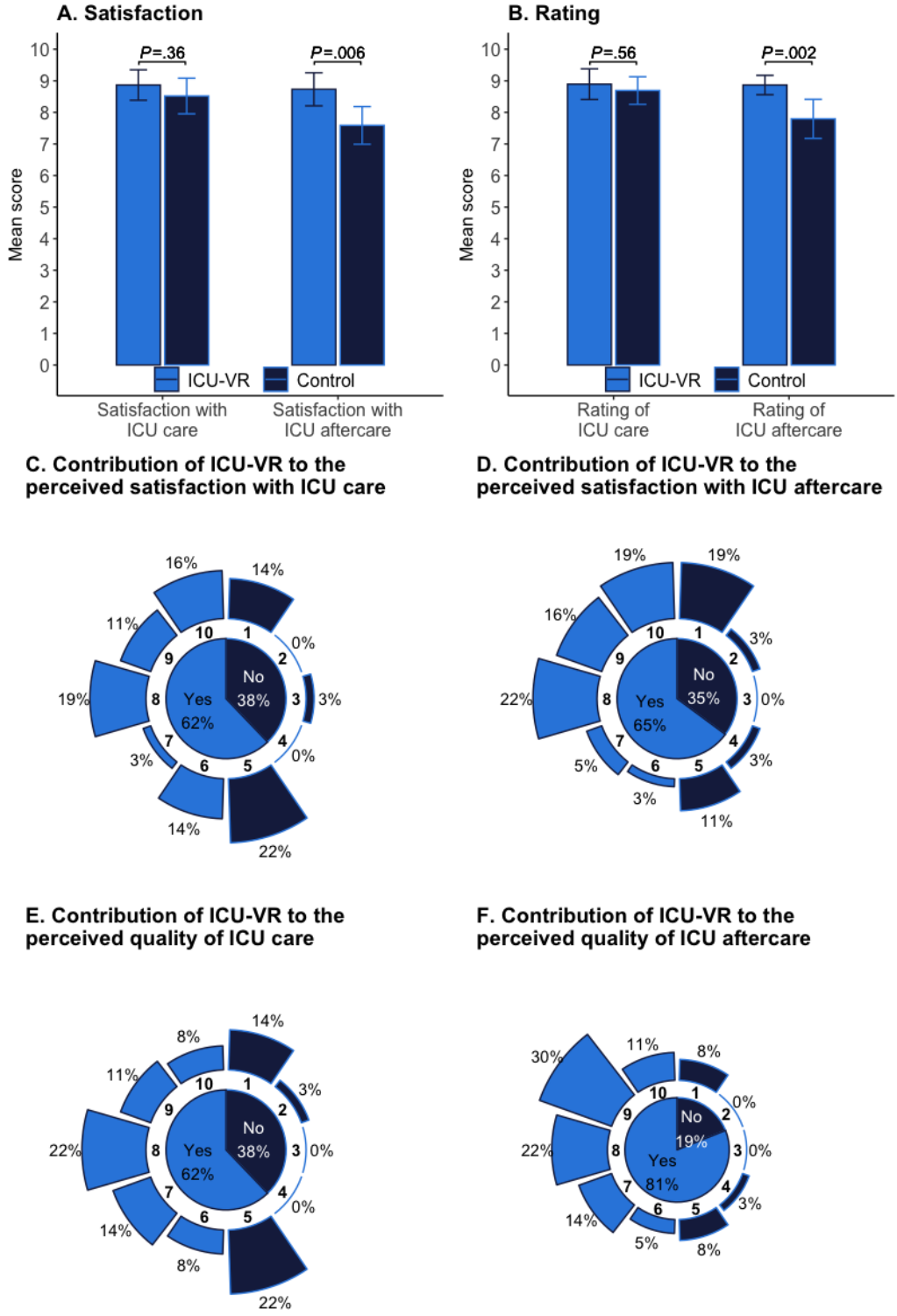


Perspectives on ICU-VR

In total, 37 patients (84%) in the ICU-VR group and 32 patients (71%) in the control group gave their perspective about the intervention and the received care and aftercare (Figure 4 and Multimedia Appendix 4**).** Patients in the intervention group were more satisfied with the ICU aftercare (mean score: 8.7, SD 1.6 vs 7.6, SD 1.6 [ICU-VR vs control], t_{64}=–2.82, *P*=.006) but not with the ICU care (mean score 8.9, SD 1.5 vs 8.5, SD 1.5 [ICU-VR vs control], t_{64}=–0.92; *P*=.36; Figure 4A). Additionally, patients in the intervention group rated the ICU aftercare higher (mean overall rating of aftercare 8.9, SD 0.9 vs 7.8, SD 1.7 [ICU-VR vs control], t_{64}=–3.25; *P*=.002) but not the ICU care (mean score 8.9, SD 1.5 vs 8.7, SD 1.2 [ICU-VR vs control], t_{64}=–0.59; *P*=.56; Figure 4B). ICU-VR added to the satisfaction of ICU care according to 62% of patients (Figure 4C), satisfaction with ICU aftercare according to 65% of patients (Figure 4D), quality of ICU care according to 62% patients (Figure 4E), and quality of ICU aftercare according to 81% of patients in the ICU-VR group (Figure 4F).

Patients in the intervention group assigned a mean score of 8.7 (SD 1.0) out of 10 to the ICU-VR, on a Likert scale, and stated that ICU-VR improved their understanding of ICU treatment (score>5, 76%; mean score 7.2, SD 2.5) and decreased their frightening memories (score>5, n=24-37, 65%; mean score 6.6, SD 2.8; Multimedia Appendix 4).

Figure 4. Perspectives on ICU-VR. Bar charts of the mean satisfaction score (A) and rating (B) of ICU care (left) and ICU aftercare (right) in the ICU-VR and control group, wherein the error bars indicate the 95% CI of the scores. The contribution of ICU-VR to the perceived satisfaction with ICU care (C) and ICU aftercare (D) and the contribution of ICU-VR to the perceived quality of ICU care (E) and ICU aftercare (F) are presented as combined pie/bar charts, indicating the percentage of patients in the ICU-VR group who gave a score above 5 (inner circle) and the percentage of patients in the ICU-VR group giving a certain score (outer circle). ICU: intensive care unit, ICU-VR: intensive care unit–virtual reality.



Discussion

Principal Findings

We observed that ICU-VR improved patients' perceived quality of, satisfaction with, and rating of ICU aftercare among COVID-19 ICU survivors. This method is feasible, acceptable, and innovative and could be implemented in regional ICU aftercare. Our results also demonstrate that approximately 31% of COVID-19 ICU survivors experienced decreased mental health in terms of psychological distress up to 6 months after

hospital discharge and that ICU-VR did not improve psychological recovery or quality of life**.**

In contrast to our previous findings regarding patients with sepsis and a recent COVID-19 case report, we did not observe improved mental health or quality of life in the ICU-VR group [19,20]. In contrast to this study, we provided ICU-VR earlier post ICU admission (median 7-8 days) in our previous study. Notably, the patient with COVID-19 in our case report and those in the sepsis study had robust responses in terms of psychological distress symptoms, including PTSD. The COVID-19 critical illness survivors in this study received ICU-VR much later (3 months after hospital discharge). Therefore, the timing of the ICU-VR intervention could be important for its therapeutic effect. Although 3 months after hospital discharge is a clinically feasible time point, it can be argued that PTSD and anxiety, at that moment, have already fully developed, and treatment of fully established psychiatric disorders may require more complex treatment strategies. When ICU-VR is offered soon after ICU discharge; that is, in the initial few weeks patients are still processing what happened to them, and ICU-VR could be a valuable adjunct to improve factual recall and decrease frightening memories. In future studies, the timing of ICU-VR and the number of sessions needed should be further investigated.

The number of desired or needed VR sessions remains a matter of debate, and no study has determined the optimal number of sessions after ICU admission. Although an average of 8-14 sessions is used in nonhospital settings, we previously demonstrated that sepsis survivors desire a median of only one session [20,46,47]. In this study, more than half of the patients did not desire the ICU-VR intervention multiple times, although there was substantial interpatient variability. An important difference between the current study and the sepsis trial is that in the sepsis trial, patients could self-determine how many sessions they desired, and this could have potentially increased the effectiveness. Therefore, a more patient-centered approach instead of a prespecified number of times might be more suitable, though guidelines are currently lacking.

Additionally, we observed lower overall incidence rates of PTSD (22% vs 11%), anxiety (46% vs 21%), and depression (41% vs 18%) at 3 months compared to a recent nationwide study in the United Kingdom, which included all patients who received at least 24 hours of ICU treatment, and compared to previously observed studies involving patients with COVID-19, acute lung injury (and acute respiratory distress syndrome) survivors, and a Dutch cohort of critical illness survivors [8,9,16,48-50]. Importantly, our power calculation was based on the prevalence rates of psychological distress. This lower incidence might explain the lack of ICU-VR effectiveness for this item. The lack of predisposing factors (such as a pre-existing cognitive impairment) could possibly explain the low prevalence of PTSD and depression in the current population [51-53].

Satisfaction during and after ICU admission is increasingly becoming an issue of interest considering that low satisfaction negatively impacts psychological sequelae after critical care [54]. Evidence suggests that patients generally indicate that they are satisfied with ICU care [55,56]. We found similarly high levels of reported satisfaction and showed that despite these high numbers, ICU-VR improved satisfaction, ratings, and perceived quality of ICU aftercare. Moreover, 100% would recommend ICU-VR to other patients. This seems to suggest that satisfaction with patient care does not imply that there are no problems regarding some aspects of their inpatient experience or that they fully comprehended all ICU-related information. Our findings actually confirm the results from a recent review, which concluded that patients' support after ICU admission is multifaceted and varies across several transition points after ICU discharge [57]. An analogy can be made to civil aviation, where satisfaction may be high, but customers still complain about specific aspects of the service [58]. ICU-VR could therefore serve as an additional modality to fulfill several individual patient needs during the transition from ICU to home. Additionally, despite the lack of a successful effect on "traditional" measurements, such as the psychological questionnaires in this study, more than half of the patients experienced a decrease in frightening memories. Therefore, ICU aftercare might be more complex than we thought and may require a more patient-centered approach for measuring the results of novel intervention methods.

Limitations

Several study limitations should be acknowledged. First, despite our randomization procedure, there were statistically significant differences in primary outcome measures between groups at the 3-month follow-up time point. To ensure that no effect was overestimated, we adjusted our outcomes for the 3-month follow-up time point outcomes by adding them as independent predictors to our regression models. Although this difference was unexpected considering the randomization procedure, we could have prevented these differences by stratifying the randomization procedure on the presence of psychological distress at the follow-up time point prior to randomization. In future studies, we should consider this when possible. Second, as both the ICU-VR intervention and the questionnaire were in Dutch, we could only include patients able to understand the Dutch language. This may have resulted in selection bias, and we do not know how ICU-VR performs in nonnative Dutch speakers or if a translated version has an effect in these patients. This is especially of interest as, owing to language restrictions, these patients are expected to understand less of their ICU treatment than native Dutch patients and may therefore benefit more from such an intervention. Third, we used a novel set of questionnaires to assess patient experiences. Although these were based on and altered from the Patient Satisfaction Questionnaire and Family Satisfaction with ICU Care tools, these questionnaires have not yet been validated [43-45].

Conclusions

In conclusion, ICU-VR is a feasible and acceptable innovative method to improve patient satisfaction with and rating of ICU aftercare and adds to its perceived quality. We observed a low prevalence of psychological distress after COVID-19 ICU treatment, and ICU-VR did not improve psychological recovery or quality of life**.** Future studies should explore ICU-VR's widespread availability and application during ICU follow-up

and should determine whether the timing of ICU-VR impacts its effect on psychological PICS-related sequelae.

Acknowledgments

This study was supported by DSW, Stichting Theia, Stichting SGS, and BeterKeten. The funding sources had no role in the design of the study and collection, analysis, and interpretation of data or in writing the manuscript.

Authors' Contributions

JHV, JvB, EW, DG, and MEvG conceived the study and initiated the study design. MEvG is the coordinating investigator and grant holder. DG is the principal investigator. TIMK provided statistical expertise in the clinical trial design, and JHV and TIMK devised the statistical analysis plan. JvB, EW, JAML, and AFCS are the local principal investigators at each study site. MEH initiated the regional post–COVID-19 follow-up clinic. JHV and MEH composed the questionnaires used in the study. MvB and LLHS assisted in participant inclusion and data collection. MEvG and TIMK independently verified the data. JHV, JB, and MEvG wrote the first manuscript draft, and all authors helped to further draft the manuscript. All authors read and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Translation of the novel questionnaire assessing ICU care and aftercare satisfaction and global rating, and the perspectives on the ICU-specific virtual reality (ICU-VR) intervention.
[PDF File (Adobe PDF File), 401 KB - jmir_v24i1e32368_app1.pdf]

Multimedia Appendix 2
Table S1. Outcomes of the individual European Quality of Life, 5 dimensions domains.
[PDF File (Adobe PDF File), 562 KB - jmir_v24i1e32368_app2.pdf]

Multimedia Appendix 3
Table S2. Subscales of the Short-Form 36 throughout follow-up.
[PDF File (Adobe PDF File), 545 KB - jmir_v24i1e32368_app3.pdf]

Multimedia Appendix 4
Table S3. Perspectives on the ICU-specific virtual reality intervention.
[PDF File (Adobe PDF File), 529 KB - jmir_v24i1e32368_app4.pdf]

Multimedia Appendix 5
CONSORT (Consolidated Standards of Reporting Trials) checklist.
[PDF File (Adobe PDF File), 636 KB - jmir_v24i1e32368_app5.pdf]

References

1. Cuthbertson BH, Roughton S, Jenkinson D, Maclennan G, Vale L. Quality of life in the five years after intensive care: a cohort study. Crit Care 2010;14(1):R6 [FREE Full text] [doi: 10.1186/cc8848] [Medline: 20089197]
2. Lilly CM, Swami S, Liu X, Riker RR, Badawi O. Five-Year Trends of Critical Care Practice and Outcomes. Chest 2017 Oct;152(4):723-735. [doi: 10.1016/j.chest.2017.06.050] [Medline: 28800866]
3. Oeyen SG, Vandijck DM, Benoit DD, Annemans L, Decruyenaere JM. Quality of life after intensive care: a systematic review of the literature. Crit Care Med 2010 Dec;38(12):2386-2400. [doi: 10.1097/CCM.0b013e3181f3dec5] [Medline: 20838335]
4. Needham DM, Davidson J, Cohen H, Hopkins RO, Weinert C, Wunsch H, et al. Improving long-term outcomes after discharge from intensive care unit: report from a stakeholders' conference. Crit Care Med 2012 Feb;40(2):502-509. [doi: 10.1097/CCM.0b013e318232da75] [Medline: 21946660]
5. Kerckhoffs MC, Kosasi FFL, Soliman IW, van Delden JJM, Cremer OL, de Lange DW, et al. Determinants of self-reported unacceptable outcome of intensive care treatment 1 year after discharge. Intensive Care Med 2019 Jun;45(6):806-814 [FREE Full text] [doi: 10.1007/s00134-019-05583-4] [Medline: 30840124]
6. Jackson JC, Pandharipande PP, Girard TD, Brummel NE, Thompson JL, Hughes CG, Bringing to light the Risk Factors And Incidence of Neuropsychological dysfunction in ICU survivors (BRAIN-ICU) study investigators. Depression,

post-traumatic stress disorder, and functional disability in survivors of critical illness in the BRAIN-ICU study: a longitudinal cohort study. Lancet Respir Med 2014 May;2(5):369-379 [FREE Full text] [doi: 10.1016/S2213-2600(14)70051-7] [Medline: 24815803]
7. Marra A, Pandharipande PP, Girard TD, Patel MB, Hughes CG, Jackson JC, et al. Co-Occurrence of Post-Intensive Care Syndrome Problems Among 406 Survivors of Critical Illness. Crit Care Med 2018 Sep;46(9):1393-1401 [FREE Full text] [doi: 10.1097/CCM.0000000000003218] [Medline: 29787415]
8. Bienvenu OJ, Friedman LA, Colantuoni E, Dinglas VD, Sepulveda KA, Mendez-Tellez P, et al. Psychiatric symptoms after acute respiratory distress syndrome: a 5-year longitudinal study. Intensive Care Med 2018 Jan;44(1):38-47 [FREE Full text] [doi: 10.1007/s00134-017-5009-4] [Medline: 29279973]
9. Parker AM, Sricharoenchai T, Raparla S, Schneck KW, Bienvenu OJ, Needham DM. Posttraumatic stress disorder in critical illness survivors: a metaanalysis. Crit Care Med 2015 May;43(5):1121-1129. [doi: 10.1097/CCM.0000000000000882] [Medline: 25654178]
10. Jaffri A, Jaffri UA. Post-Intensive care syndrome and COVID-19: crisis after a crisis? Heart Lung 2020;49(6):883-884 [FREE Full text] [doi: 10.1016/j.hrtlng.2020.06.006] [Medline: 32690219]
11. Simpson R, Robinson L. Rehabilitation After Critical Illness in People With COVID-19 Infection. Am J Phys Med Rehabil 2020 Jun;99(6):470-474 [FREE Full text] [doi: 10.1097/PHM.0000000000001443] [Medline: 32282359]
12. Haines KJ, McPeake J, Hibbert E, Boehm LM, Aparanji K, Bakhru RN, et al. Enablers and Barriers to Implementing ICU Follow-Up Clinics and Peer Support Groups Following Critical Illness: The Thrive Collaboratives. Crit Care Med 2019 Sep;47(9):1194-1200 [FREE Full text] [doi: 10.1097/CCM.0000000000003818] [Medline: 31241499]
13. Loudon I. The principle of referral: the gatekeeping role of the GP. Br J Gen Pract 2008 Feb;58(547):128-30; discussion 130 [FREE Full text] [doi: 10.3399/bjgp08X277113] [Medline: 18307864]
14. Schofield-Robinson OJ, Lewis SR, Smith AF, McPeake J, Alderson P. Follow-up services for improving long-term outcomes in intensive care unit (ICU) survivors. Cochrane Database Syst Rev 2018 Nov 02;11:CD012701 [FREE Full text] [doi: 10.1002/14651858.CD012701.pub2] [Medline: 30388297]
15. Hofhuis JGM, Spronk PE, van Stel HF, Schrijvers AJP, Rommes JH, Bakker J. Experiences of critically ill patients in the ICU. Intensive Crit Care Nurs 2008 Oct;24(5):300-313. [doi: 10.1016/j.iccn.2008.03.004] [Medline: 18472265]
16. Vlake JH, van Genderen ME, Schut A, Verkade M, Wils E, Gommers D, et al. Patients suffering from psychological impairments following critical illness are in need of information. J Intensive Care 2020;8:6 [FREE Full text] [doi: 10.1186/s40560-019-0422-0] [Medline: 31938546]
17. Prescott HC, Girard TD. Recovery From Severe COVID-19: Leveraging the Lessons of Survival From Sepsis. JAMA 2020 Aug 25;324(8):739-740. [doi: 10.1001/jama.2020.14103] [Medline: 32777028]
18. Vlake J, Wils E, van Bommel J, Korevaar T, Gommers D, van Genderen ME. Virtual Reality Tailored to the Needs of Post-ICU Patients: A Safety and Immersiveness Study in Healthy Volunteers. Crit Care Explor 2021 May;3(5):e0388 [FREE Full text] [doi: 10.1097/CCE.0000000000000388] [Medline: 34079940]
19. Vlake JH, van Bommel J, Hellemons ME, Wils E, Gommers D, van Genderen ME. Intensive Care Unit-Specific Virtual Reality for Psychological Recovery After ICU Treatment for COVID-19; A Brief Case Report. Front Med (Lausanne) 2020;7:629086 [FREE Full text] [doi: 10.3389/fmed.2020.629086] [Medline: 33614677]
20. Vlake J, Van Bommel J, Wils EJ, Korevaar T, Bienvenu OJ, Klijn E, et al. Virtual Reality to Improve Sequelae of the Postintensive Care Syndrome: A Multicenter, Randomized Controlled Feasibility Study. Crit Care Explor 2021 Sep;3(9):e0538 [FREE Full text] [doi: 10.1097/CCE.0000000000000538] [Medline: 34549192]
21. Vlake JH, Van Bommel J, Wils E, Korevaar TIM, Hellemons ME, Schut AFC, et al. Effect of intensive care unit-specific virtual reality (ICU-VR) to improve psychological well-being and quality of life in COVID-19 ICU survivors: a study protocol for a multicentre, randomized controlled trial. Trials 2021 May 05;22(1):328 [FREE Full text] [doi: 10.1186/s13063-021-05271-z] [Medline: 33952318]
22. Vlake JH, Van Bommel J, Wils E, Korevaar TIM, Bienvenu OJ, Klijn E, et al. Virtual Reality to Improve Sequelae of the Postintensive Care Syndrome: A Multicenter, Randomized Controlled Feasibility Study. Crit Care Explor 2021 Sep;3(9):e0538 [FREE Full text] [doi: 10.1097/CCE.0000000000000538] [Medline: 34549192]
23. Vlake H. Intensive Care Unit-specific Virtual Reality for COVID-19 survivors. YouTube. 2020. URL: https://www.youtube.com/watch?v=Z9Mf-esnz94 [accessed 2021-11-11]
24. Weiss D. The Impact of Event Scale: Revised. In: Cross-Cultural Assessment of Psychological Trauma and PTSD. Boston, MA: Springer; 2007:219-238.
25. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand 1983 Jun;67(6):361-370. [doi: 10.1111/j.1600-0447.1983.tb09716.x] [Medline: 6880820]
26. Bienvenu OJ, Williams JB, Yang A, Hopkins RO, Needham DM. Posttraumatic stress disorder in survivors of acute lung injury: evaluating the Impact of Event Scale-Revised. Chest 2013 Jul;144(1):24-31 [FREE Full text] [doi: 10.1378/chest.12-0908] [Medline: 23699588]
27. Chan KS, Aronson Friedman L, Bienvenu OJ, Dinglas VD, Cuthbertson BH, Porter R, et al. Distribution-based estimates of minimal important difference for hospital anxiety and depression scale and impact of event scale-revised in survivors

of acute respiratory failure. Gen Hosp Psychiatry 2016;42:32-35 [FREE Full text] [doi: 10.1016/j.genhosppsych.2016.07.004] [Medline: 27638969]
28. Creamer M, Bell R, Failla S. Psychometric properties of the Impact of Event Scale - Revised. Behav Res Ther 2003 Dec;41(12):1489-1496. [doi: 10.1016/j.brat.2003.07.010] [Medline: 14705607]
29. Bjelland I, Dahl AA, Haug TT, Neckelmann D. The validity of the Hospital Anxiety and Depression Scale. Journal of Psychosomatic Research 2002 Feb;52(2):69-77. [doi: 10.1016/s0022-3999(01)00296-3] [Medline: 11832252]
30. Sukantarat KT, Williamson RCN, Brett SJ. Psychological assessment of ICU survivors: a comparison between the Hospital Anxiety and Depression scale and the Depression, Anxiety and Stress scale. Anaesthesia 2007 Mar;62(3):239-243 [FREE Full text] [doi: 10.1111/j.1365-2044.2006.04948.x] [Medline: 17300300]
31. Jutte JE, Needham DM, Pfoh ER, Bienvenu OJ. Psychometric evaluation of the Hospital Anxiety and Depression Scale 3 months after acute lung injury. J Crit Care 2015 Aug;30(4):793-798 [FREE Full text] [doi: 10.1016/j.jcrc.2015.04.006] [Medline: 25981443]
32. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
33. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9] [Medline: 10109801]
34. Angus DC, Carlet J, 2002 Brussels Roundtable Participants. Surviving intensive care: a report from the 2002 Brussels Roundtable. Intensive Care Med 2003 Mar;29(3):368-377. [doi: 10.1007/s00134-002-1624-8] [Medline: 12536269]
35. Granja C, Teixeira-Pinto A, Costa-Pereira A. Quality of life after intensive care--evaluation with EQ-5D questionnaire. Intensive Care Med 2002 Jul;28(7):898-907. [doi: 10.1007/s00134-002-1345-z] [Medline: 12122528]
36. Badia X, Diaz-Prieto A, Gorriz MT, Herdman M, Torrado H, Farrero E, et al. Using the EuroQol-5D to measure changes in quality of life 12 months after discharge from an intensive care unit. Intensive Care Med 2001 Dec;27(12):1901-1907. [doi: 10.1007/s00134-001-1137-x] [Medline: 11797026]
37. Heyland DK, Hopman W, Coo H, Tranmer J, McColl MA. Long-term health-related quality of life in survivors of sepsis. Short Form 36: a valid and reliable measure of health-related quality of life. Crit Care Med 2000 Nov;28(11):3599-3605. [doi: 10.1097/00003246-200011000-00006] [Medline: 11098960]
38. Chrispin PS, Scotton H, Rogers J, Lloyd D, Ridley SA. Short Form 36 in the intensive care unit: assessment of acceptability, reliability and validity of the questionnaire. Anaesthesia 1997 Jan;52(1):15-23 [FREE Full text] [doi: 10.1111/j.1365-2044.1997.015-az014.x] [Medline: 9014540]
39. M Versteegh M, M Vermeulen K, M A A Evers S, de Wit GA, Prenger R, A Stolk E. Dutch Tariff for the Five-Level Version of EQ-5D. Value Health 2016 Jun;19(4):343-352 [FREE Full text] [doi: 10.1016/j.jval.2016.01.003] [Medline: 27325326]
40. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. Health Econ 1993 Oct;2(3):217-227. [doi: 10.1002/hec.4730020305] [Medline: 8275167]
41. Ware J, Kosinski M, Keller S. SF-36 Physical and Mental Health Summary Scales: A User's Manual. Boston, MA: ScienceOpen; 1994.
42. Aaronson NK, Muller M, Cohen PD, Essink-Bot ML, Fekkes M, Sanderman R, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. J Clin Epidemiol 1998 Nov;51(11):1055-1068. [doi: 10.1016/s0895-4356(98)00097-3] [Medline: 9817123]
43. Rensen A, van Mol MM, Menheere I, Nijkamp MD, Verhoogt E, Maris B, et al. Quality of care in the intensive care unit from the perspective of patient's relatives: development and psychometric evaluation of the consumer quality index 'R-ICU'. BMC Health Serv Res 2017 Jan 24;17(1):77 [FREE Full text] [doi: 10.1186/s12913-016-1975-4] [Medline: 28118840]
44. van Mol MMC, Bakker EC, Nijkamp MD, Kompanje EJO, Bakker J, Verharen L. Relatives' perspectives on the quality of care in an Intensive Care Unit: the theoretical concept of a new tool. Patient Educ Couns 2014 Jun;95(3):406-413. [doi: 10.1016/j.pec.2014.03.019] [Medline: 24746926]
45. Marshall G, Hays R. The patient satisfaction questionnaire short-form (PSQ-18). Rand. 1994. URL: https://www.rand.org/content/dam/rand/pubs/papers/2006/P7865.pdf [accessed 2022-01-10]
46. Wechsler TF, Kümpers F, Mühlberger A. Inferiority or Even Superiority of Virtual Reality Exposure Therapy in Phobias?-A Systematic Review and Quantitative Meta-Analysis on Randomized Controlled Trials Specifically Comparing the Efficacy of Virtual Reality Exposure to Gold Standard Exposure in Agoraphobia, Specific Phobia, and Social Phobia. Front Psychol 2019;10:1758 [FREE Full text] [doi: 10.3389/fpsyg.2019.01758] [Medline: 31551840]
47. Kothgassner OD, Goreis A, Kafka JX, Van Eickels RL, Plener PL, Felnhofer A. Virtual reality exposure therapy for posttraumatic stress disorder (PTSD): a meta-analysis. Eur J Psychotraumatol 2019;10(1):1654782 [FREE Full text] [doi: 10.1080/20008198.2019.1654782] [Medline: 31489138]
48. Rogers JP, Chesney E, Oliver D, Pollak TA, McGuire P, Fusar-Poli P, et al. Psychiatric and neuropsychiatric presentations associated with severe coronavirus infections: a systematic review and meta-analysis with comparison to the COVID-19 pandemic. Lancet Psychiatry 2020 Jul;7(7):611-627 [FREE Full text] [doi: 10.1016/S2215-0366(20)30203-0] [Medline: 32437679]

49. Nikayin S, Rabiee A, Hashem MD, Huang M, Bienvenu OJ, Turnbull AE, et al. Anxiety symptoms in survivors of critical illness: a systematic review and meta-analysis. Gen Hosp Psychiatry 2016;43:23-29 [FREE Full text] [doi: 10.1016/j.genhosppsych.2016.08.005] [Medline: 27796253]
50. Rabiee A, Nikayin S, Hashem MD, Huang M, Dinglas VD, Bienvenu OJ, et al. Depressive Symptoms After Critical Illness: A Systematic Review and Meta-Analysis. Crit Care Med 2016 Sep;44(9):1744-1753 [FREE Full text] [doi: 10.1097/CCM.0000000000001811] [Medline: 27153046]
51. Pinkas J, Horowitz A. Reducing Severity of Posttraumatic Stress Disorder in Intensive Care Unit Survivors. Dimens Crit Care Nurs 2020;39(6):298-304. [doi: 10.1097/DCC.0000000000000439] [Medline: 33009267]
52. Fernandez-Gonzalo S, Turon M, De Haro C, López-Aguilar J, Jodar M, Blanch L. Do sedation and analgesia contribute to long-term cognitive dysfunction in critical care survivors? Med Intensiva (Engl Ed) 2018 Mar;42(2):114-128. [doi: 10.1016/j.medin.2017.06.010] [Medline: 28851588]
53. Van Rompaey B, Elseviers MM, Schuurmans MJ, Shortridge-Baggett LM, Truijen S, Bossaert L. Risk factors for delirium in intensive care patients: a prospective cohort study. Crit Care 2009;13(3):R77 [FREE Full text] [doi: 10.1186/cc7892] [Medline: 19457226]
54. Kavalnienė R, Deksnyte A, Kasiulevičius V, Šapoka V, Aranauskas R, Aranauskas L. Patient satisfaction with primary healthcare services: are there any links with patients' symptoms of anxiety and depression? BMC Fam Pract 2018 Jun 19;19(1):90 [FREE Full text] [doi: 10.1186/s12875-018-0780-z] [Medline: 29921234]
55. Mukhopadhyay A, Song G, Sim PZ, Ting KC, Yoo JKS, Wang QL, et al. Satisfaction Domains Differ between the Patient and Their Family in Adult Intensive Care Units. Biomed Res Int 2016;2016:9025643 [FREE Full text] [doi: 10.1155/2016/9025643] [Medline: 28044138]
56. Holanda Peña MS, Talledo NM, Ots Ruiz E, Lanza Gómez JM, Ruiz Ruiz A, García Miguelez A, Proyecto HU-CI. Satisfaction in the Intensive Care Unit (ICU). Patient opinion as a cornerstone. Med Intensiva 2017 Mar;41(2):78-85 [FREE Full text] [doi: 10.1016/j.medin.2016.06.007] [Medline: 27793389]
57. King J, O'Neill B, Ramsay P, Linden MA, Darweish Medniuk A, Outtrim J, et al. Identifying patients' support needs following critical illness: a scoping review of the qualitative literature. Crit Care 2019 May 24;23(1):187 [FREE Full text] [doi: 10.1186/s13054-019-2441-6] [Medline: 31126335]
58. Bethune G. From Worst to First: Behind the Scenes of Continental's Remarkable Comeback. Hoboken, NJ: John Wiley & Sons; 1998.

Abbreviations

EDC: electronic data capture
EQ-5D: European Quality of Life, 5 dimensions
HADS: Hospital Anxiety and Depression Scale
HRQoL: health-related quality of life
ICU: intensive care unit
ICU-VR: intensive care unit–specific virtual reality
IES-R: Impact of Event Scale-Revised
OR: odds ratio
PICS: postintensive care syndrome
PTSD: posttraumatic stress disorder
RT–PCR: reverse transcription–polymerase chain reaction
SF-36: Short-Form 36
VR: virtual reality

Edited by G Eysenbach; submitted 09.08.21; peer-reviewed by S Ouyang, T Ong; comments to author 10.11.21; revised version received 01.12.21; accepted 29.12.21; published 31.01.22.

Please cite as:
Vlake JH, van Bommel J, Wils EJ, Bienvenu J, Hellemons ME, Korevaar TIM, Schut AFC, Labout JAM, Schreuder LLH, van Bavel MP, Gommers D, van Genderen ME
Intensive Care Unit–Specific Virtual Reality for Critically Ill Patients With COVID-19: Multicenter Randomized Controlled Trial
J Med Internet Res 2022;24(1):e32368
URL: https://www.jmir.org/2022/1/e32368
doi:10.2196/32368
PMID:34978530

Original Paper

Online Patient Education Materials Related to Lipoprotein(a): Readability Assessment

Keon Pearson[1*], MD; Summer Ngo[2*], BSc; Eson Ekpo[1], MD; Ashish Sarraju[2], MD; Grayson Baird[3], PhD; Joshua Knowles[2], MD, PhD; Fatima Rodriguez[2], MPH, MD

[1]Department of Medicine, Stanford University, Stanford, CA, United States

[2]Division of Cardiovascular Medicine and the Cardiovascular Institute, Stanford University, Stanford, CA, United States

[3]Rhode Island Hospital and Diagnostic Imaging, Warren Alpert Medical School, Brown University, Providence, RI, United States

[*]these authors contributed equally

Corresponding Author:
Keon Pearson, MD
Department of Medicine
Stanford University
300 Pasteur Drive
Room S102
Stanford, CA, 94305
United States
Phone: 1 6507236028
Email: keonp@stanford.edu

Abstract

Background: Lipoprotein(a) (Lp(a)) is a highly proatherogenic lipid fraction that is a clinically significant risk modifier. Patients wanting to learn more about Lp(a) are likely to use online patient educational materials (OPEMs). However, the readability of OPEMs may exceed the health literacy of the public.

Objective: This study aims to assess the readability of OPEMs related to Lp(a). We hypothesized that the readability of these online materials would exceed the sixth grade level recommended by the American Medical Association.

Methods: Using an online search engine, we queried the top 20 search results from 10 commonly used Lp(a)-related search terms to identify a total of 200 websites. We excluded duplicate websites, advertised results, research journal articles, or non–patient-directed materials, such as those intended only for health professionals or researchers. Grade level readability was calculated using 5 standard readability metrics (automated readability index, SMOG index, Coleman-Liau index, Gunning Fog score, Flesch-Kincaid score) to produce robust point (mean) and interval (CI) estimates of readability. Generalized estimating equations were used to model grade level readability by each search term, with the 5 readability scores nested within each OPEM.

Results: A total of 27 unique websites were identified for analysis. The average readability score for the aggregated results was a 12.2 (95% CI 10.9798-13.3978) grade level. OPEMs were grouped into 6 categories by primary source: industry, lay press, research foundation and nonprofit organizations, university or government, clinic, and other. The most readable category was OPEMs published by universities or government agencies (9.0, 95% CI 6.8-11.3). The least readable OPEMs on average were the ones published by the lay press (13.0, 95% CI 11.2-14.8). All categories exceeded the sixth grade reading level recommended by the American Medical Association.

Conclusions: Lack of access to readable OPEMs may disproportionately affect patients with low health literacy. Ensuring that online content is understandable by broad audiences is a necessary component of increasing the impact of novel therapeutics and recommendations regarding Lp(a).

(J Med Internet Res 2022;24(1):e31284) doi:10.2196/31284

KEYWORDS

lipoprotein(a); readability; online patient education material; health education; health literacy

Introduction

Lipoprotein(a) (Lp(a)) is a highly proatherogenic lipid fraction that is increasingly recognized as a clinically significant risk modifier. The 2018 American College of Cardiology and the American Heart Association guidelines recommend using Lp(a) as a risk-enhancing factor favoring initiation of statin therapy in persons with intermediate atherosclerotic cardiovascular disease (ASCVD) risk [1]. This recommendation is based on data in recent years showing a strong genetic determination of Lp(a) levels; limited modifiability with diet, exercise, and medication; and an increased risk of ASCVD independent of traditional risk factors [2]. National cohort studies using Mendelian randomization estimate that reducing Lp(a) by 50 mg/dl and 99 mg/dl could reduce major adverse cardiac events by 20% and 40%, respectively [3]. Statins do not lower Lp(a), and there are currently no Food and Drug Administration–approved therapies to lower Lp(a) specifically. PCSK-9 inhibitors result in a modest reduction in Lp(a) and may be associated with a reduction in major adverse cardiovascular events for patients with elevated baseline Lp(a) who have experienced acute coronary syndrome [4,5]. Antisense oligonucleotides targeting expression of Lp(a) are in phase III clinical trials [6,7]. The National Lipid Association published guidance in 2019 for clinicians using Lp(a) in clinical practice, and as novel therapeutics are approved, interest in this biomarker is expected to increase in the coming years [8]. Patients may be especially interested in learning their Lp(a) levels after encountering stories of public figures experiencing early ASCVD in the setting of an elevated Lp(a) [9].

Patients wanting to learn more about Lp(a) are likely to use online patient educational materials (OPEMs) [10]. However, patient health literacy may influence utilization of evidence-based OPEMs [11,12]. The American Medical Association (AMA) recommends writing health information for patients at the sixth grade level or below to ensure broad comprehension [13]. Although OPEMs influence patient decision-making, the readability of OPEMs generally exceeds the health literacy of the public [14-16]. Patients with lower health literacy have been noted to have poorer overall health and higher mortality, and this is notably true as well among patients from racial and ethnic minority backgrounds, partially explaining the racial disparities in some outcomes [17]. The disparate impact of elevated Lp(a) in racial/ethnic groups and the lack of standardization of the use of this biomarker in clinical practice make it even more pressing to ensure that evidence-based materials are appropriately written for the public. Thus, we sought to quantify the readability of frequently accessed OPEMs about Lp(a).

Methods

Data Acquisition and Refinement

We used Google, the largest online search engine, to query the first 20 results for each of the following 10 search terms: “lipoprotein(a),” “lipoprotein(a) cardiovascular risk,” “lipoprotein(a) elevated,” “lipoprotein(a) high,” “lipoprotein(a) levels,” “lipoprotein(a) screening,” “lipoprotein(a) test,” “Lp little a,” “Lp(a),” and “Lp(a) screening” [18]. Location, cookies, and user account information were disabled beforehand to avoid search bias. All 200 websites were accessed and downloaded as PDFs on November 5 to 6, 2020. OPEMs were defined as materials intended for patients and the public. Two independent reviewers reviewed each source’s mission statement or informational page to determine whether it was patient-facing. Research journal articles, advertised results, results failing to contain material on Lp(a), and non–patient-directed sources such as those intended only for health professionals or researchers were excluded.

This research is exempt from human participant institutional review board approval as no human participant data were used, and only publicly available OPEMs were included in the analysis.

OPEMs were grouped into 6 categories by primary source. *Industry OPEMs* were published by for-profit companies or offered a proprietary test or service related to Lp(a). *Lay press OPEMs* were published by news organizations and health care reporters that do not have a specific research or scientific focus. *Research foundations and nonprofit OPEMs* were published by health-related organizations that have a specific research or scientific focus. *University or government OPEMs* were published by academic institutions or national, state, or local government agencies. *Clinic OPEMs* were published by Williams Integracare clinic, an independently owned family medicine, chiropractic, and physical therapy clinic in St. Cloud, MN. *Other* includes Wikipedia.org, a crowdsourced online free encyclopedia, which published an article on Lp(a).

Readability Assessment

Websites meeting the criteria for OPEMs were converted into plain text in separate Word (Microsoft Corporation) documents and prepared in accordance with recommendations from the Centers for Medicare and Medicaid Services prior to scoring readability [19]. Using methods consistent with prior readability assessment studies we removed advertisements, videos, images, figures, captions, hyperlinks, disclaimers, copyright notices, acknowledgments, and citations. Periods were used to denote the end of all sentences; all other punctuation were removed. Symbols and numerals were spelled out to avoid artificial increases in reading level. Readability was assessed with Readable.com, as done in prior literature [16].

Statistical Methods

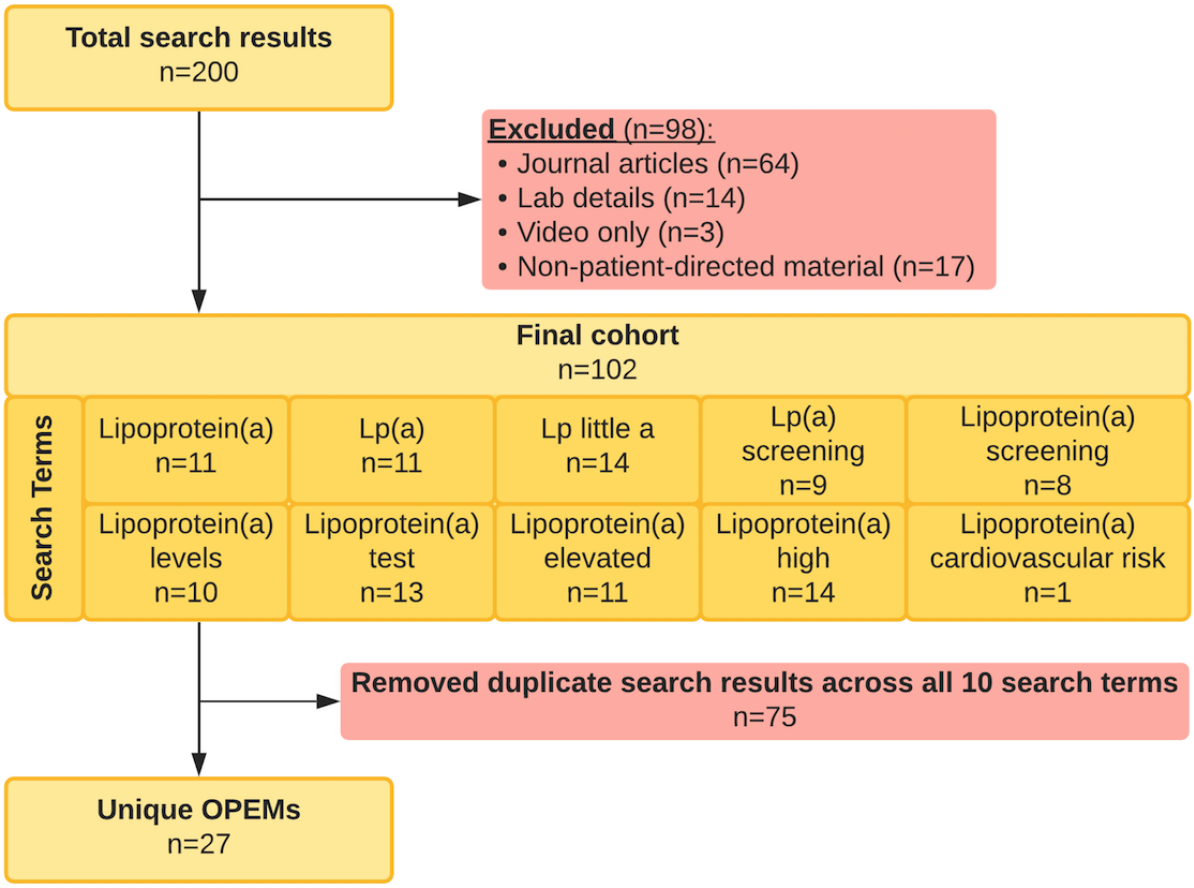
Similar to prior readability studies, grade-level readability was calculated using 5 standard readability metrics (automated readability index, SMOG index, Coleman-Liau index, Gunning Fog score, Flesch-Kincaid score) to produce robust point (mean) and interval (CI) estimates of readability. Averaging across multiple readability metrics has been demonstrated to yield more reliable results than relying on a single readability metric [20]. This was done to minimize bias because no single readability metric has been established as a “gold standard”; each readability metric is calculated differently and varies in limitations [21]. Generalized estimating equations were used to model grade-level readability by each search term, with the 5 readability scores nested within each OPEM [16]. Readability

analyses were conducted with SAS Software 9.4 (SAS Institute), with sandwich estimation and the GLIMMIX procedure. All interval estimates were calculated for 95% confidence. The interval estimates reflect the variability of readability for the 5 readability metrics for each OPEM.

Results

Figure 1 summarizes the selection Lp(a) OPEMs for analysis. From an initial sample of 200 total search results across all 10 search terms, 102 results met inclusion criteria. A total of 75 duplicate results were then removed. There were 27 unique OPEMs included in the readability analysis.

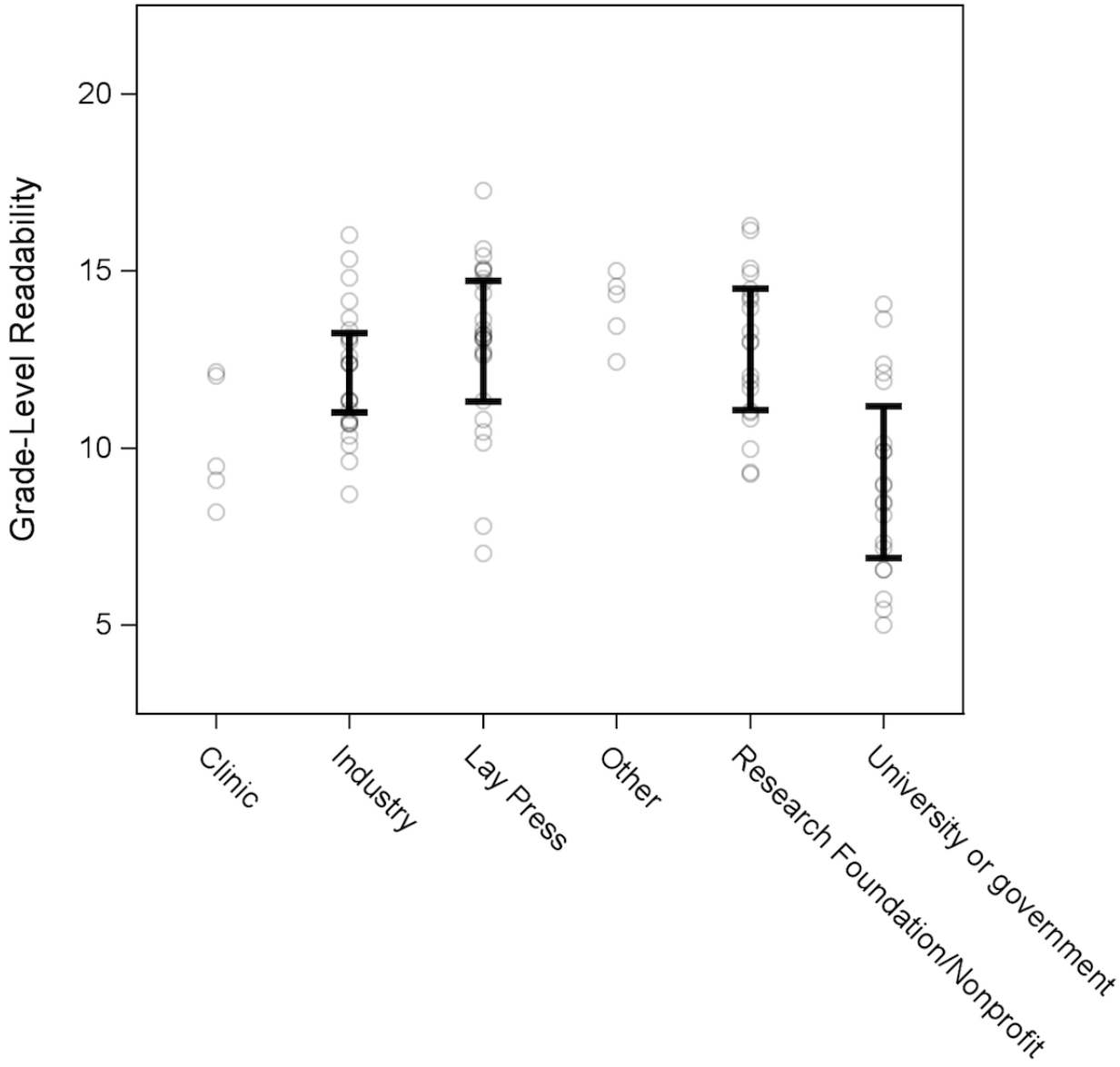
Figure 1. Material selection and exclusion for OPEMs related to lipoprotein(a). OPEM: online patient education material.



Among the 27 OPEMs, the largest category of were university and government sources, which included 30% (n=8) of unique search results. The next largest categories were research and nonprofit foundations and industry publications, which each comprised 22% (n=6) of the results. The third largest category of OPEMs were articles in the lay press, which comprised 19% (n=5/27) of results. The smallest categories were clinic publications and other, both of which had 1 (3%) website (Williams Integracare Clinic and Wikipedia, respectively).

The average readability score across unique websites was at a 12.2 (95% CI 10.9798-13.3978) grade level. The most readable category was OPEMs published by universities or government agencies (9.0, 95% CI 6.8-11.3). The least readable OPEMs on average were the ones published by lay press (13.0, 95% CI 11.2-14.8). Research and nonprofit foundations (12.8, 95% CI 11.0-14.6) and industry (12.1, 95% CI 10.9-13.3) had intermediate readability. Clinic (10.2) and other publications (14) had 1 site per category and therefore are not averaged figures. Figure 2 summarizes the readability scores by category of OPEMs.

Figure 2. Average grade level readability of online patient education material (OPEM) by category of publication. Each circle represents a readability score for one OPEM, with a total of five readability scores for each unique OPEM. 95% CIs are included for all categories except "Clinic" and "Other," which only had one unique OPEM each. University and government sources were significantly more readable than research and nonprofit foundation, industry, and lay press sources.



Of the 27 unique search results, only 1 site, the University of Rochester Medical Center (7.0, 95% CI 5.7-8.3), had the sixth grade level recommended by the AMA within its 95% CI. The least readable OPEM was the HEART UK's Lp(a) general information site (15.7, 95% CI 14.8-16.6). The readability of each unique website is shown in Figure 3.

Figure 3. Readability rankings by search result. Each blue dot corresponds to the mean readability based on the average of five standard readability scores (automated readability index, SMOG index, Coleman-Liau index, Gunning fog score, Flesch-Kincaid score) with whiskers representing the range of readability scores. The red vertical line corresponds to the AMA sixth grade level readability target. Cleveland Clinic1: Elevated Lipoprotein(a): Is a Long-Sought Treatment Finally on the Way?; Cleveland Clinic2: Why Would My Doctor Order a Lipoprotein(a) Blood Test?; National Lipid Association1: lipoprotein(a) Screening for Individuals at High ASCVD Risk; National Lipid Association2: Elevated Lipoprotein (a) Patient-Centered Education From the National Lipid Association. AMA: American Medical Association.

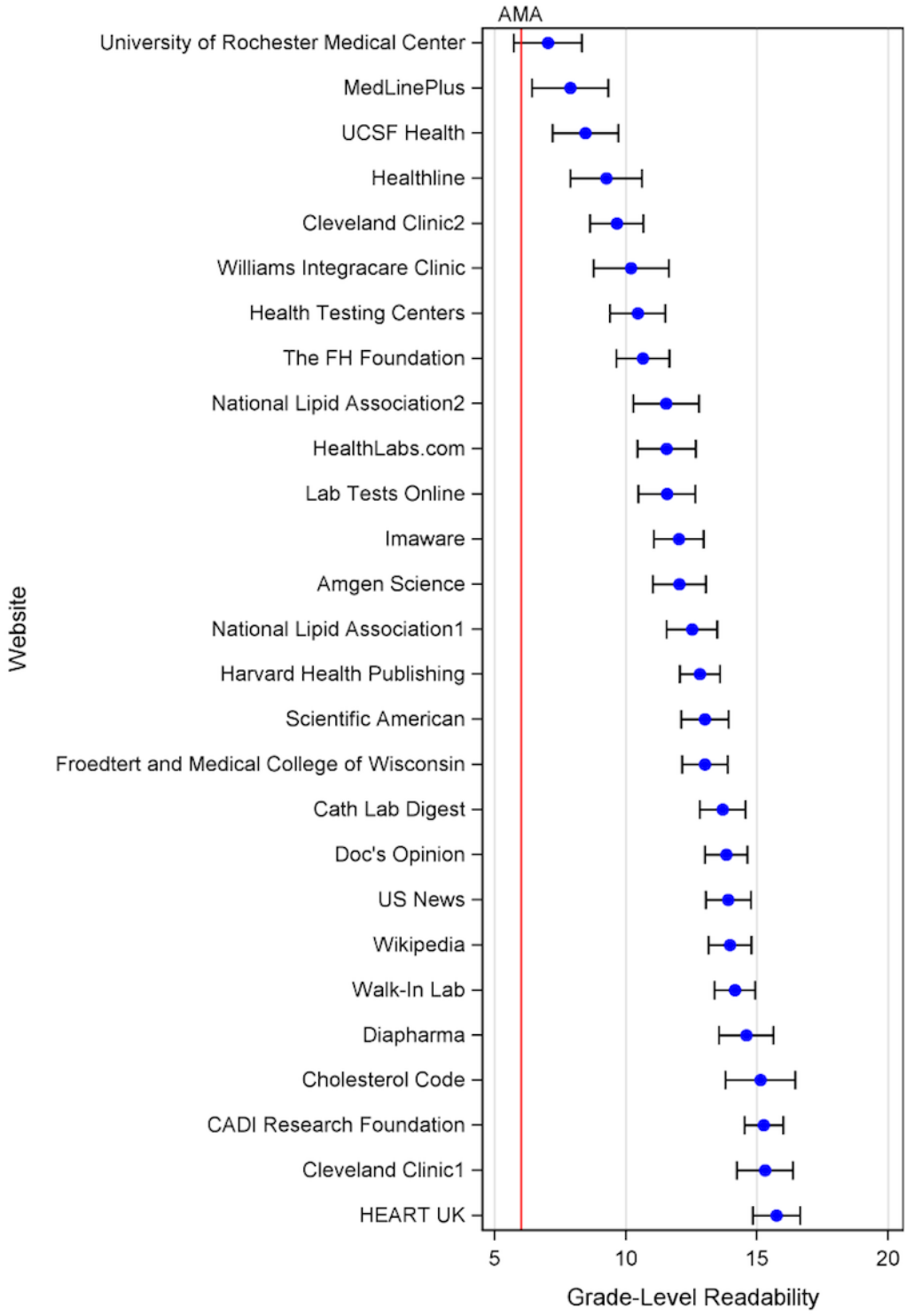


Table 1 highlights excerpts from Lp(a) OPEMs across readability levels.

Table 1. Search concepts and excerpts of readable and nonreadable quotes from OPEM.

Concept	Excerpt from less readable OPEM[a]	Excerpt from more readable OPEM
Lp(a)[b] contains both a lipid and a protein carrier	"Lipoprotein(a), or Lp(a), is a distinctive particle with two components: a lipoprotein core that resembles LDL, along with a shell that contains apolipoprotein(a), or apo(a)." [22]	"Lipoproteins are substances made up of protein and fat" [23]
Lp(a) is a proatherogenic lipoprotein fraction	"High levels of LP(a) have now been identified as an independent risk factor in cardiovascular disease, with a causal link to atherosclerosis (furring up of arteries), heart attacks, strokes, aortic valve disease and heart failure." [24]	"High levels of Lp(a) can create plaque in your blood vessels." [25]
High-risk populations should be screened for Lp(a)	"Measurement of lipoprotein(a) is now recommended in several patient subgroups… patients with premature atherosclerosis; patients with a strong family history of premature coronary heart disease (CHD); patients with elevated LDL-C and greater than or equal to two risk factors; patients who have had coronary angioplasty in whom lipoprotein(a) excess may increase the risk of restenosis; patients who have undergone coronary bypass graft surgery in whom Lp(a) excess may be associated with graft stenosis." [26]	"You may need this test if you have: Heart disease, despite normal results on other lipid tests, High cholesterol, despite maintaining a healthy diet A family history of heart disease, especially heart disease that has occurred at an early age and/or sudden deaths from heart disease." [27]
There is no widely implemented standard for measuring Lp(a)	"Although the reference material for the accurate measurement of Lp(a) … has been available for many years, many commercial laboratories have not changed their reagents and testing methods and continue to use old reagents and methods resulting in inaccurate results. Accordingly, results of Lp(a) measurements by different labs are not comparable and some of them are clearly inaccurate." [28]	"Note: Normal value ranges may vary slightly among different laboratories. Talk to your doctor about the meaning of your specific test results." [29]
There are no drugs demonstrated to improve outcomes in patients with elevated Lp(a)	"Results using statin medications have been mixed in most trials … In severe cases, such as familial hypercholesterolemia or treatment- resistant hypercholesterolemia, lipid apheresis may dramatically reduce Lp(a) … Other medications that are in various stages of development include thyromimetics, cholesterol-ester-transfer protein (CETP inhibitors), anti-sense oligonucleopeptides, and proprotein convertase subtilisin/kexin type 9 (PCSK-9) inhibitors." [30]	"Medications/treatments in current use that lower Lp(a) also lower cholesterol. There are apheresis and niacin. These both have substantial side effects. PCSK9 inhibitors lower Lp(a) while lowering LDL cholesterol. Statins have no effect on Lp(a)." [31]

[a]OPEM: online patient educational material.

[b]Lp(a): lipoprotein(a).

A list of the OPEM sources used for this study can be found in Multimedia Appendix 1.

Discussion

Principal Results

We found that the average reading grade level of OPEMs pertaining to Lp(a) generally exceeded AMA readability recommendations that OPEMs be written at or below a sixth grade reading level to be accessible to the public. Average grade-level readability of OPEMs on Lp(a) was a 12.2 (95% CI 10.9798-13.3978) grade level, exceeding the average reading level of US adults (eighth grade) [13]. Of 27 unique websites reviewed, only 1 website had a lower bound of reading grade level (5.7) that was below the sixth grade reading level recommendation. Our results suggest that the overwhelming majority of Lp(a) OPEMs are written at a reading level that is too high for the minimally health literate members of the public.

These findings have several important implications for how patients may make decisions about this important, proatherogenic lipid fraction that is receiving increased attention. Patients frequently use the internet to supplement health information from their clinicians [32]. If presented clearly, online health information can be a valuable patient resource. Patients who feel more informed are more comfortable asking their provider questions and report better understanding of their providers' explanations and greater self-confidence in making health care decisions [33]. A 2018 cross-sectional survey found that patients who searched online for health information to solve their medical problems were also significantly more likely to change their medical decision based on information gathered [34]. It is noteworthy that OPEMs in the academic and government categories were the most readable. Academic and government sites are regarded as reliable sources given that many organizations seek to advance public health and knowledge [35]. Given the major influence of online health information on decision-making, these academic and government websites may also benefit from direct guidance to craft OPEMs with readability targets in view. Despite the importance of these OPEMs in reaching the lay community, research foundations and nonprofit sites demonstrate the largest gaps between patient reading skills and OPEMs reading level on Lp(a).

The less readable OPEMs tend to cover topics in greater depth, including nuances around Lp(a) measurement and standardization, and the role of Lp(a) in thrombogenesis and wound healing. Thus, there appears to be a trade-off between readability and comprehensiveness in OPEMs.

These findings align with results from prior studies across a broad range of health conditions that show OPEMs commonly exceed the recommended readability level. Ayyaswami et al [36] showed that greater than 99% of OPEMs relating to cardiovascular disease were written above the grade level recommended by the AMA [35]. OPEMs are frequently written at a reading level too difficult for the public to comprehend, and low readability levels have been documented across disciplines including common topics related to surgery, oncology, and radiology [37,38].

The percentage of adults with below basic health literacy is considerably higher for populations who identify as Black (24%), Hispanic (41%), American Indian/Alaska Native (25%), and Asian/Pacific Islander (13%) compared to non-Hispanic White (9%) [39]. Similarly, over half of adults older than 65 years were found to have less than a basic health literacy level [39]. These are the very populations known to face a disproportionate burden of cardiovascular risk. Our current risk prediction models such as the Pooled Cohort Equations do not adequately capture the risk to heterogenous racial/ethnic groups [1,40-42]. The importance of reaching these populations is further increased given that Lp(a) is known to circulate at higher levels in patients of African and South Asian descent [43,44].

Many clinical preventive and screening services are underused by historically marginalized racial/ethnic communities and older adults due to inadequate health care access and low health literacy, but we suggest that providing more readable OPEMs may help bridge this gap in care. In the interim, our study findings remind clinicians to consider the readability of OPEMs and patient literacy when recommending Lp(a) evaluation. Shared decision-making requires adequate understanding of the risks and benefits of any diagnostic testing or risk stratification procedure. Actionability of OPEMs on Lp(a) is presently limited by the lack of approved therapies; however, with changes in emerging therapies and practice guidelines, understandability and actionability will be important parameters to assess systematically using the Patient Education Materials Assessment Tool in future studies.

A strength of our study is that we incorporated readability results from five different standard readability metrics, which allows us to have a robust evaluation of readability regardless of the number of websites evaluated. All of the reading grade level estimates supported our hypothesis that Lp(a) OPEMs are written above the recommended sixth grade reading level. We also included a thorough review of possible patient queries by analyzing 200 search results for 10 commonly used search terms. Finally, we included examples of communication of a concept from both more readable and less readable OPEMs. This may serve as a real-world, practical guide for creators of OPEMs seeking to choose words and phrases that will be accessible to a broad audience.

Limitations

Our study should be interpreted in the context of certain limitations. We did not account for other search engines besides Google. This limitation is somewhat mitigated by the fact that 88% of global internet users use Google as their most frequent search engine [18]. As with other OPEM readability studies, the readability metrics used here do not consider the inherent complexity of some medical terms. Polysyllabic words and longer words are automatically rated as more complex and less understandable than short or monosyllabic words, which in medicine does not always hold true.

Conclusions

In conclusion, we found that the grade level readability of OPEMs relating to Lp(a) generally substantially exceeded the sixth grade reading level recommended by the AMA. This gap in readability may disproportionately affect patients with low health literacy. Creators of OPEMs should be mindful of the readability of their content. Ensuring that online content is understandable by broad audiences is a necessary component of increasing the impact of novel therapeutics and recommendations regarding Lp(a).

Acknowledgments

The authors of the study thank the Family Heart Foundation for assistance in reviewing the early drafts of this paper.

Authors' Contributions

KP, SN, FR, and GB designed the study. GB performed all statistical analysis. SN performed all searches and filtered the online patient education materials, as well as conducted the initial readability analysis. KP and AS reviewed the websites for patient-centeredness. KP and SN wrote the manuscript. EE designed all figures and tables. FR, JK, and AS provided editorial review.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Included online patient educational materials URLs.
[DOCX File , 15 KB - jmir_v24i1e31284_app1.docx]

References

1. Grundy SM, Stone NJ, Bailey AL, Beam C, Birtcher KK, Blumenthal RS, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA Guideline on the Management of Blood Cholesterol: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. Circulation 2019 Jun 18;139(25):e1082-e1143 [FREE Full text] [doi: 10.1161/CIR.0000000000000625] [Medline: 30586774]
2. Pearson K, Rodriguez F. Lipoprotein(a) and cardiovascular disease prevention across diverse populations. Cardiol Ther 2020 Dec;9(2):275-292 [FREE Full text] [doi: 10.1007/s40119-020-00177-4] [Medline: 32451810]
3. Madsen CM, Kamstrup PR, Langsted A, Varbo A, Nordestgaard BG. Lipoprotein(a)-lowering by 50 mg/dL (105 nmol/L) may be needed to reduce cardiovascular disease 20% in secondary prevention: a population-based study. Arterioscler Thromb Vasc Biol 2020 Jan;40(1):255-266 [FREE Full text] [doi: 10.1161/ATVBAHA.119.312951] [Medline: 31578080]
4. O'Donoghue ML, Fazio S, Giugliano RP, Stroes ESG, Kanevsky E, Gouni-Berthold I, et al. Lipoprotein(a), PCSK9 inhibition, and cardiovascular risk. Circulation 2019 Mar 19;139(12):1483-1492. [doi: 10.1161/CIRCULATIONAHA.118.037184] [Medline: 30586750]
5. Bittner VA, Szarek M, Aylward PE, Bhatt DL, Diaz R, Edelberg JM, ODYSSEY OUTCOMES Committees and Investigators. Effect of alirocumab on lipoprotein(a) and cardiovascular risk after acute coronary syndrome. J Am Coll Cardiol 2020 Jan 21;75(2):133-144 [FREE Full text] [doi: 10.1016/j.jacc.2019.10.057] [Medline: 31948641]
6. Santos RD, Raal FJ, Catapano AL, Witztum JL, Steinhagen-Thiessen E, Tsimikas S. Mipomersen, an antisense oligonucleotide to apolipoprotein B-100, reduces lipoprotein(a) in various populations with hypercholesterolemia: results of 4 phase III trials. Arterioscler Thromb Vasc Biol 2015 Mar;35(3):689-699 [FREE Full text] [doi: 10.1161/ATVBAHA.114.304549] [Medline: 25614280]
7. Tsimikas S, Viney NJ, Hughes SG, Singleton W, Graham MJ, Baker BF, et al. Antisense therapy targeting apolipoprotein(a): a randomised, double-blind, placebo-controlled phase 1 study. Lancet 2015 Oct 10;386(10002):1472-1483. [doi: 10.1016/S0140-6736(15)61252-1] [Medline: 26210642]
8. Wilson DP, Jacobson TA, Jones PH, Koschinsky ML, McNeal CJ, Nordestgaard BG, et al. Use of Lipoprotein(a) in clinical practice: a biomarker whose time has come. A scientific statement from the National Lipid Association. J Clin Lipidol 2019;13(3):374-392. [doi: 10.1016/j.jacl.2019.04.010] [Medline: 31147269]
9. O'Connor A. A heart risk factor even doctors know little about. The New York Times. 2018 Jan 9. URL: https://www.nytimes.com/2018/01/09/well/heart-risk-doctors-lipoprotein.html [accessed 2021-03-07]
10. Rice RE. Influences, usage, and outcomes of Internet health information searching: multivariate results from the Pew surveys. Int J Med Inform 2006 Jan;75(1):8-28. [doi: 10.1016/j.ijmedinf.2005.07.032] [Medline: 16125453]
11. Ilic D. The role of the internet on patient knowledge management, education, and decision-making. Telemed J E Health 2010;16(6):664-669. [doi: 10.1089/tmj.2010.0003] [Medline: 20575610]
12. Hersh L, Salzman B, Snyderman D. Health literacy in primary care practice. Am Fam Physician 2015 Jul 15;92(2):118-124 [FREE Full text] [Medline: 26176370]
13. Weiss B. Health Literacy and Patient Safety: Help Patients Understand: Manual for Clinicians Internet. Chicago, IL: AMA Foundation; 2007.
14. Imoisili OE, Levinsohn E, Pan C, Howell BA, Streiter S, Rosenbaum JR. Discrepancy between patient health literacy levels and readability of patient education materials from an electronic health record. Health Lit Res Pract 2017 Oct;1(4):e203-e207 [FREE Full text] [doi: 10.3928/24748307-20170918-01] [Medline: 31294265]
15. Kapoor K, George P, Evans MC, Miller WJ, Liu SS. Health literacy: readability of ACC/AHA online patient education material. Cardiology 2017;138(1):36-40. [doi: 10.1159/000475881] [Medline: 28571004]
16. Rodriguez F, Ngo S, Baird G, Balla S, Miles R, Garg M. Readability of online patient educational materials for coronary artery calcium scans and implications for health disparities. J Am Heart Assoc 2020 Sep 15;9(18):e017372 [FREE Full text] [doi: 10.1161/JAHA.120.017372] [Medline: 32865121]
17. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med 2011 Jul 19;155(2):97-107. [doi: 10.7326/0003-4819-155-2-201107190-00005] [Medline: 21768583]
18. Worldwide desktop market share of leading search engines from January 2010 to September 2021. Statista. URL: https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/ [accessed 2021-11-24]
19. Toolkit for making written material clear and effective: section 4: special topics for writing and design: part 7: using readability formulas: a cautionary note. Centers for Medicare and Medicaid Services. 2010 Sep 30. URL: https://www.cms.gov/Outreach-and-Education/Outreach/WrittenMaterialsToolkit/Downloads/ToolkitPart07.pdf [accessed 2021-11-25]
20. Mailloux SL, Johnson ME, Fisher DG, Pettibone TJ. How reliable is computerized assessment of readability? Comput Nurs 1995;13(5):221-225. [Medline: 7585304]
21. Zhou S, Jeong H, Green PA. How consistent are the best-known readability equations in estimating the readability of design standards? IEEE Trans Professional Commun 2017 Mar;60(1):97-111. [doi: 10.1109/tpc.2016.2635720]
22. Elevated lipoprotein(a): is a long-sought treatment finally on the way? Consult QD. URL: https://consultqd.clevelandclinic.org/elevated-lipoproteina-is-a-long-sought-treatment-finally-on-the-way/ [accessed 2021-11-25]
23. Lipoprotein-A test. Healthline. URL: https://www.healthline.com/health/lipoprotein-a [accessed 2021-11-25]

24. High lipoprotein (a). HEART UK. URL: https://www.heartuk.org.uk/genetic-conditions/high-lipoproteina [accessed 2021-11-25]
25. Lipoprotein(a) cholesterol. University of Rochester Medical Center. URL: https://www.urmc.rochester.edu/encyclopedia/content.aspx?contenttypeid=167&contentid=lpa_cholesterol#:~:text=Lp(a)%20is%20a%20type,%2Ddensity%20lipoprotein%20(LDL).&text=A%20high%20level%20of%20Lp,risk%20factor%20for%20heart%20disease [accessed 2021-11-25]
26. Lipoprotein (a) blood test. Walk-In Lab. URL: https://www.walkinlab.com/products/view/lipoprotein-a-blood-test [accessed 2021-11-25]
27. Lipoprotein (a) blood test. MedlinePlus. URL: https://medlineplus.gov/lab-tests/lipoprotein-a-blood-test/ [accessed 2021-11-25]
28. Lp (a) testing and standardization. Cadi. URL: https://cadiresearch.org/topic/lipoproteina/lpa-testing-2 [accessed 2021-11-25]
29. Lipoprotein-a. UCSF Health. URL: https://www.ucsfhealth.org/medical-tests/lipoprotein-a [accessed 2021-11-25]
30. Lipoprotein(a). Wikipedia. URL: https://en.wikipedia.org/wiki/Lipoprotein(a) [accessed 2021-11-25]
31. Lp(a) and familial hypercholesterolemia. FH Foundation. URL: https://thefhfoundation.org/lpa-and-familial-hypercholesterolemia#:~:text=What%20is%20Lp(a)%3F,levels%20than%20the%20general%20population [accessed 2021-11-25]
32. Lagan BM, Sinclair M, Kernohan WG. Internet use in pregnancy informs women's decision making: a web-based survey. Birth 2010 Jun;37(2):106-115. [doi: 10.1111/j.1523-536X.2010.00390.x] [Medline: 20557533]
33. Empowerment and Engagement Among Low-income Californians: Enhancing Patient-Centered Care. San Francisco, CA: Blue Shield of California Foundation; 2021:E.
34. Chen Y, Li C, Liang J, Tsai C. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. J Med Internet Res 2018 Feb 12;20(2):e47 [FREE Full text] [doi: 10.2196/jmir.9370] [Medline: 29434017]
35. Bergeron B. Online patient-education options. MedGenMed 2004 Apr 28;6(2):54 [FREE Full text] [Medline: 15266279]
36. Ayyaswami V, Padmanabhan D, Patel M, Prabhu AV, Hansberry DR, Agarwal N, et al. A readability analysis of online cardiovascular disease-related health education materials. Health Lit Res Pract 2019 Apr;3(2):e74-e80 [FREE Full text] [doi: 10.3928/24748307-20190306-03] [Medline: 31049489]
37. Miles RC, Choi P, Baird GL, Dibble EH, Lamb L, Garg M, et al. Will the effect of new federal breast density legislation be diminished by currently available online patient educational materials? Acad Radiol 2020 Oct;27(10):1400-1405. [doi: 10.1016/j.acra.2019.11.008] [Medline: 31839567]
38. Miles RC, Baird GL, Choi P, Falomo E, Dibble EH, Garg M. Readability of online patient educational materials related to breast lesions requiring surgery. Radiology 2019 Apr;291(1):112-118. [doi: 10.1148/radiol.2019182082] [Medline: 30694156]
39. White S, Bennett I, Cordell T, Baxter S. Health literacy of America's adults: results from the 2003 National Assessment of Adult Literacy. PsycEXTRA Dataset 2006 Sep 06:A. [doi: 10.1037/e530912012-001]
40. Tota-Maharaj R, Blaha MJ, Blankstein R, Silverman MG, Eng J, Shaw LJ, et al. Association of coronary artery calcium and coronary heart disease events in young and elderly participants in the multi-ethnic study of atherosclerosis: a secondary analysis of a prospective, population-based cohort. Mayo Clin Proc 2014 Oct;89(10):1350-1359 [FREE Full text] [doi: 10.1016/j.mayocp.2014.05.017] [Medline: 25236430]
41. Goff DC, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, American College of Cardiology/American Heart Association Task Force on Practice Guidelines. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. Circulation 2014 Jun 24;129(25 Suppl 2):S49-S73 [FREE Full text] [doi: 10.1161/01.cir.0000437741.48606.98] [Medline: 24222018]
42. Rodriguez F, Chung S, Blum MR, Coulet A, Basu S, Palaniappan LP. Atherosclerotic cardiovascular disease risk prediction in disaggregated Asian and Hispanic subgroups using electronic health records. J Am Heart Assoc 2019 Jul 16;8(14):e011874 [FREE Full text] [doi: 10.1161/JAHA.118.011874] [Medline: 31291803]
43. Enas EA, Varkey B, Dharmarajan TS, Pare G, Bahl VK. Lipoprotein(a): an underrecognized genetic risk factor for malignant coronary artery disease in young Indians. Indian Heart J 2019;71(3):184-198 [FREE Full text] [doi: 10.1016/j.ihj.2019.04.007] [Medline: 31543191]
44. Virani SS, Brautbar A, Davis BC, Nambi V, Hoogeveen RC, Sharrett AR, et al. Associations between lipoprotein(a) levels and cardiovascular outcomes in black and white subjects: the Atherosclerosis Risk in Communities (ARIC) Study. Circulation 2012 Jan 17;125(2):241-249 [FREE Full text] [doi: 10.1161/CIRCULATIONAHA.111.045120] [Medline: 22128224]

Abbreviations

AMA: American Medical Association
ASCVD: atherosclerotic cardiovascular disease
Lp(a): lipoprotein(a)
OPEM: online patient educational material

Edited by R Kukafka; submitted 16.06.21; peer-reviewed by E Neter, S Rush; comments to author 14.09.21; revised version received 25.10.21; accepted 29.10.21; published 11.01.22.

Please cite as:
Pearson K, Ngo S, Ekpo E, Sarraju A, Baird G, Knowles J, Rodriguez F
Online Patient Education Materials Related to Lipoprotein(a): Readability Assessment
J Med Internet Res 2022;24(1):e31284
URL: https://www.jmir.org/2022/1/e31284
doi:10.2196/31284
PMID:35014955

Original Paper

Optimizing Readability and Format of Plain Language Summaries for Medical Research Articles: Cross-sectional Survey Study

Leia Martínez Silvagnoli[1], MSc; Caroline Shepherd[2], BPharm; James Pritchett[1], PhD; Jason Gardner[2], PhD

[1]Department of Life Sciences, Faculty of Science and Engineering, Manchester Metropolitan University, Manchester, United Kingdom

[2]CMC Connect, McCann Health Medical Communications, Macclesfield, United Kingdom

Corresponding Author:
Caroline Shepherd, BPharm
CMC Connect
McCann Health Medical Communications
Complete House
19–21 King Edward Street
Macclesfield, SK10 1AQ
United Kingdom
Phone: 44 1625 62 4000
Email: Caroline.Shepherd@connect-cmc.com

Abstract

Background: Plain language summaries (PLSs) are intended to provide readers with a clear, nontechnical, and easily understandable overview of medical and scientific literature; however, audience preferences for specific PLS formats have yet to be fully explored.

Objective: This study aims to evaluate the preferred readability level and format for PLSs of medical research articles of different disease states via a web-based survey of audiences of different age groups.

Methods: Articles describing phase III clinical trials published in top-level, peer-reviewed journals between May 2016 and May 2018 were identified for 3 chronic disease states representing a range of adult patient age groups: (1) psoriasis, a skin disease representative of younger patients; (2) multiple sclerosis (MS), a neurological disease representative of middle-aged patients; and (3) rheumatoid arthritis (RA), a painful joint disease representative of older patients. Four PLSs were developed for each research article, of which 3 were text-only summaries (written with high, medium, and low complexity) and 1 was an infographic. To evaluate each of the 4 PLS formats, a 20-question open survey (specific to one of the 3 diseases) was sent to a representative sample selected via UK-based patient association websites, Twitter, and Facebook patient groups. A weighted-average calculation was applied to respondents' ranked preferences for each PLS format.

Results: For all 3 articles, the weighted-average preference scores showed that infographic (psoriasis 2.91, MS 2.71, and RA 2.78) and medium-complexity text-based PLS (reading age 14-17 years, US Grade 9-11; psoriasis 2.90; MS 2.47; RA 2.77) were the two most preferred PLS formats.

Conclusions: Audience preferences should be accounted for when preparing PLSs to accompany peer-reviewed original research articles. Oversimplified text can be viewed negatively, and graphical summaries or medium-complexity text-based summaries appear to be the most popular.

Plain Language Summary: Patients and caregivers should have the chance to read about medical research in a format they can understand. However, we do not know much about the formats that people with different illnesses or ages prefer. Researchers wanted to find out more about this. They selected 3 medical articles about illnesses that affect different age groups: psoriasis (younger patients), multiple sclerosis (middle-aged patients), and rheumatoid arthritis (older patients). They created 4 summaries of each article. One was a graphical summary, and the other 3 were words-only summaries of high, medium, and low complexity. Then, the researchers posted surveys on UK patient group websites and Facebook patient groups to ask people what they thought of the summaries. The surveys were taken by 167 people. These people were patients with psoriasis, multiple sclerosis, or rheumatoid arthritis, or their caregivers. Most were women, and about half had a university degree. For each illness, most people preferred the graphical summary. Among the word-only summaries, most people preferred the medium-complexity wording written for a reading age of 14 to 17 years. People felt that the graphical and medium-complexity summaries were clear and concise, while the others used jargon or were too simple. Authors of medical articles should remember these results when writing

summaries for patients. More research is needed about the preferences of other people, such as those with other illnesses. (See Multimedia Appendix 1 for the graphical summary of the plain language summary.)

(J Med Internet Res 2022;24(1):e22122) doi:10.2196/22122

KEYWORDS

biomedical research; health literacy; multiple sclerosis; plain language summary; psoriasis; rheumatoid arthritis

Introduction

Background

Health literacy, that is, the degree to which individuals can obtain, process, and understand basic health information to make appropriate health decisions [1,2], is critical to the patient-doctor relationship [3]. Health information should be easy to access, use, and understand for everyone, including both patients and their caregivers. However, despite the increasing availability of medical content from different forms of media, studies have shown that few nonexperts can understand, or act on, the health information available [2] and that text is often written above the general readability level, in a way that limits understanding and hinders the ability to make informed choices [3,4]. Indeed, in a 2019 survey of more than 14,000 people in the United States, 88% of respondents thought that "scientists should be sharing their results in easy-to-understand language" [5].

Plain Language Summaries

Plain language summaries (PLSs) have been introduced to make written and verbal information more easily understood by nonexperts [6-8]. Such strategies are gradually being adopted across all documents, presentations, and electronic communications intended for the public to avoid the use of jargon and highly technical language, and to focus on the information that is most relevant for patients, caregivers, and families [1,7]. Text is written in an easily readable style with short, clear sentences, using everyday English words, and avoiding complex grammatical structures wherever possible [6]. Thus, a PLS can explain complicated medical research to the nonexpert, thereby extending the reach of scientific information and empowering nonexperts with the knowledge to act on the information they receive [7-10].

Through the use of PLSs, scientific information is given a direct route from researchers to a broader audience beyond the scientific community. PLSs provide greater clarity to all those interested in learning about expert scientific material [11], while reducing the risk of overinterpretation via journalism or social media [12]. It is important to recognize that PLSs can be for everyone—from nonspecialists, including patients, caregivers, the lay public, and nonexperts in the field of research [1,6], to busy medical specialists and other healthcare providers [6,12]. For healthcare professionals, establishing new standards of communication, such as PLS, will improve their ability to meet the needs of quickly changing health systems and increasingly globalized populations [13]. Furthermore, a wide distribution of information is also expected to improve patient and healthcare provider engagement [1,2,10,13], promoting an increased focus on disease research and public support. Many research organizations now have public blogs on their websites, which discuss certain aspects of their research that may not necessarily be covered by scientific publications [11]. Encouraging public involvement in this way can improve the quality of research and also help with the development of new research strategies [11].

PLS is a term used to cover many forms of summary information in the medical or scientific setting. It is important to make the distinction between two of the most common forms of PLS, as explained below.

Types of PLS

The first is a clinical trial summary (CTS), where clinical trial sponsors produce a brief summary of the trial, focusing on the main results (ie, the primary endpoint and key safety data). These summary documents are shared with trial participants and the general public; they are usually posted on the sponsor's website or an independent electronic repository. CTS are a mandatory requirement of the European Union Clinical Trials Regulation and Good Lay Summary Practice (GLSP) recommendations have recently been published as part of the EudraLex Volume 10 clinical trials guidelines [14]; for the United States, a draft guidance document making similar CTS recommendations was submitted to the FDA in 2017 [1,15,16]. The elements that must be contained within a CTS are strictly defined within these regulatory guidelines. CTSs are not the subject of our research.

The second form of PLSs, which is the focus of our research, relates to summarizing a peer-reviewed article published in a medical journal [9,12]. These PLSs act as easy-to-read executive summaries of the most up-to-date research published in the medical literature. They are usually optional and published as a free, open-access document alongside the associated medical journal article. Hereafter we refer to a PLS in the context of summarizing the medical literature.

Development of PLSs

The benefits of PLSs have been recognized [17]; previous studies have aimed to understand different stakeholders' perspectives on PLSs [18]; and helpful tools are available to assist with the development of PLSs [6,19-21]. However, research on the most effective communication strategies remains limited. For example, it is unclear whether most audiences prefer text-based articles or more visual formats using infographics (ie, graphs and charts that provide clear information) [8,22,23]. Crick and Hartling [23] found that doctors preferred PLSs in text format, whereas nurses preferred an infographic format. Buljan et al [9] found that students, doctors, and consumers (female members of a patient and parent action group) reported no difference in the knowledge they obtained from infographic

or text-based PLSs. Therefore, although these studies offer interesting insights, there is little evidence regarding the preferred format of PLSs of publications read by lay audiences, considering populations representative of those seeking information from the medical literature.

Different text readability formulas are available to aid the development of PLSs [24-26], but the level of complexity that should be applied to text-based PLSs remains to be established. A survey of the adult general population in England indicated that approximately half the population has only basic literacy skills, of General Certificate of Secondary Education (GCSE) Grade D and below [27]. The UK Government Digital Service suggests that content should be developed to reflect the reading age of a 9-year-old child [28]. Furthermore, expert group recommendations for CTSs of European registered clinical trials state that these summaries should normally be accessible by young people from the age of 12 years of age and above, and that sponsors should consider testing CTS readability among those representing the target population [15]. However, it remains to be determined which literacy level(s) should be considered when developing PLSs of medical literature and whether this would differ with topic (eg, disease type) and the age range of the target reader.

Study Aim

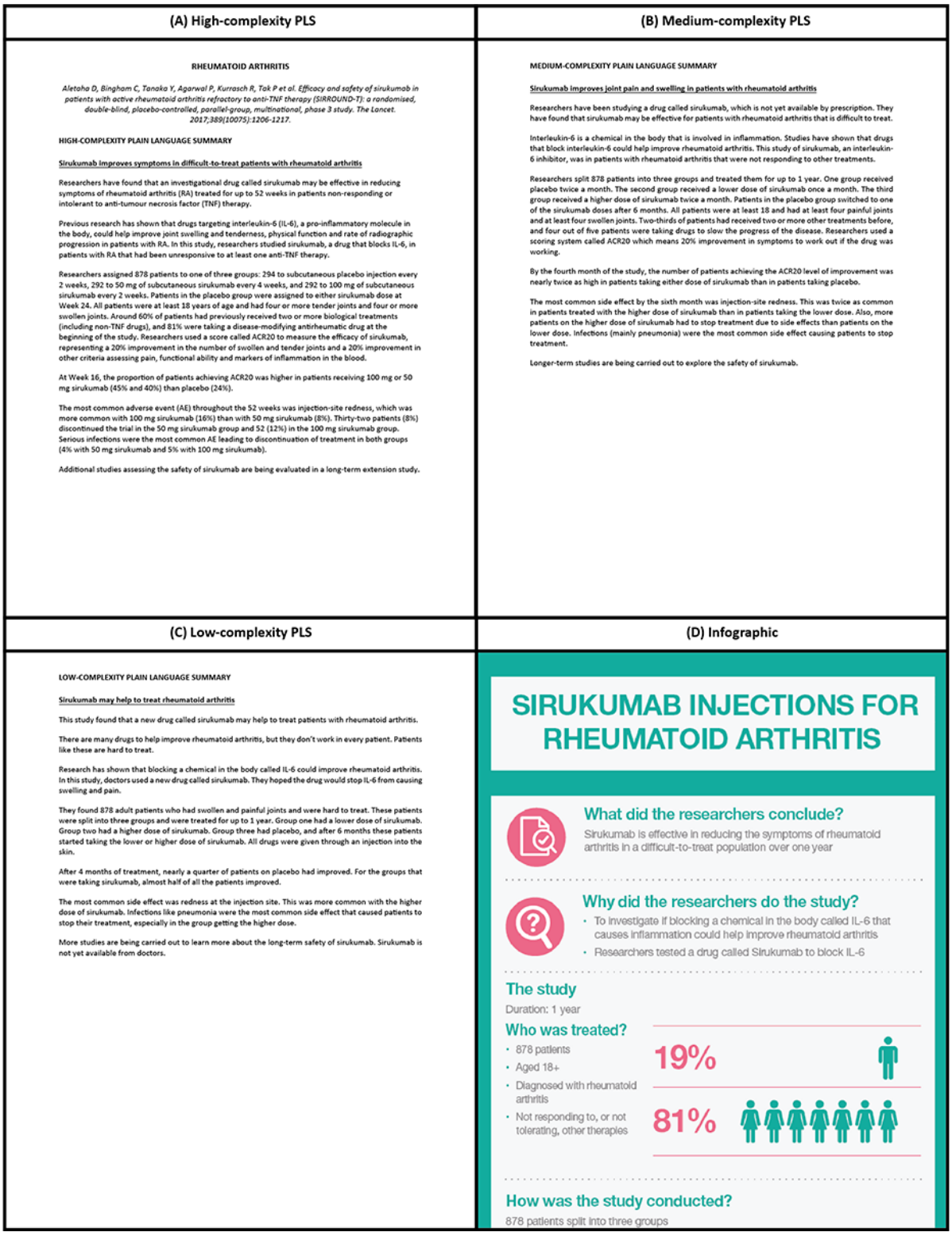
This cross-sectional study aims to evaluate the preferred readability and format for PLSs of medical research pertaining to chronic diseases affecting different age ranges, among web-based, lay audiences (ie, patients and caregivers) who may likely have an interest in obtaining information about the latest research in the field.

Methods

Three chronic diseases were chosen, representing different age band classifications based on the age groups commonly affected by these conditions—psoriasis, representative of a predominantly younger population; multiple sclerosis (MS), representative of a predominantly middle age group; and rheumatoid arthritis (RA), representative of predominantly older patients. To source relevant articles, journals were selected based on their impact factor and narrowed down to those that published research articles focusing on all 3 diseases. Specific articles were identified using the PubMed database, searching for randomized controlled phase III trials published from May 2016 to May 2018.

One article was selected for each chronic disease [29-31]. Four PLSs were developed for each of the 3 articles; of these, 3 PLSs were text-only summaries (ie, written with high, medium, and low levels of complexity) and the fourth PLS was an infographic (see Figure 1 and Multimedia Appendices 2, 3, 4). Complexity of the text (based on main body text only) was determined using an automated readability checker from the Readabilityformulas website [24]. Varying levels of complexity (ie, high, medium, and low) for the text-only PLSs were measured and adapted by changing variables such as the length and number of sentences, syllable count, and use of acronyms (a summary of the differences in text complexity for each PLS vs the abstract of each original article is provided in Multimedia Appendix 5).

Figure 1. Examples of various PLS formats used. (A) High-complexity text-only PLS, (B) medium-complexity text-only PLS, (C) low-complexity text-only PLS, and (D) infographic PLS format. Text complexity in each case was determined using an automated readability checker from Readabilityformulas website [24], using text from the main body only (ie, excluding title, authors, and funding statements) and omitting any parenthetical data. Full versions of the infographics analyzed are shown in Multimedia Appendices 1-3. PLS: plain language summary.



A 20-question (1 question per page), web-based open survey was developed using SurveyMonkey [12] to assess the readability of, and preference for, each of the 4 PLS formats (Multimedia Appendix 6). The order of presentation of the PLSs to the survey respondents was as follows: (1) high-complexity text-only PLS, (2) medium-complexity text-only PLS, (3) low-complexity text-only PLS, and (4) infographic format. The usability and technical functionality of the survey were tested by one of the authors' colleagues who had no scientific qualifications. The survey was sent to organizations representing patients and caregivers for each of the 3 conditions. The survey was accessed via UK-based patient association websites and Facebook patient support groups (Multimedia Appendix 7). Specific associations and patient support groups approached

for this study include Psoriasis and Psoriatic Arthritis Alliance, Psoriasis Association, Psoriasis Support Group UK, MS Society, Multiple Sclerosis Trust, MS-UK, Asian MS (UK national support group), Multiple Sclerosis Support/Chat Group UK, Mutual Support (Armed Forces), National Rheumatoid Arthritis Society Regional groups, Arthritis Research UK, Arthritis Care (part of Arthritis Research UK), Arthritis Action, and the UK Rheumatoid Arthritis Wonky Group. Each organization chose its wording to advertise the survey, based on the background information provided. Participation in the survey was voluntary, and questions could be skipped (no nonresponse options were provided). Respondents could move back and forward throughout the survey to review and change their answers before submission. No incentives to complete the survey were provided. The survey was active for 3 weeks, between August 10 and September 2, 2018.

Ethical approval was obtained from the Manchester Metropolitan University Research Ethics Committee. All survey responses were anonymous, and no personal information or identifying information were collected or made available to the researchers.

Participants were informed of the scholarly purpose of the study, details of the principal investigator, estimated length of time for survey completion, and anonymity of data they were to provide. Cookies inherent to the SurveyMonkey platform were used, which prevented duplicate entries [32]. Some data were not collected, including any assessment of unique site visitors, view rate (ie, ratio of unique survey visitors or unique site visitors), and participation rate (ie, ratio of unique visitors who agreed to participate or unique first survey page visitors).

The completion rate was determined by calculating the ratio of total number of respondents who finished the survey to total number of respondents who initiated the survey. All data were included in the analysis, regardless of whether the survey was fully completed.

No formal statistical analyses were performed on these data. A weighted-average calculation, performed through the SurveyMonkey platform, determined the average ranking for each PLS option to identify the most preferred format. The format with the highest average ranking score indicates the respondents' preferred option.

The average ranking was calculated as follows:

where w is the weight of the ranked position, and x is the response count for the corresponding answer choice.

For each person who responded, the most preferred choice (ranked as #1) was assigned the largest weight (in this case: 4); by contrast, the least preferred choice (ranked #4) was assigned a weight of 1. No data adjustments were made.

Subgroup analyses were also performed to identify PLS preference based on individuals' age (younger, 18-34 years; middle-aged, 35-54 years; or older, ≥55 years), gender (female, male, or other), and education level (nondegree level or university degree level, defined as including a UK university bachelor's degree, master's degree, PhD, or other postgraduate degree).

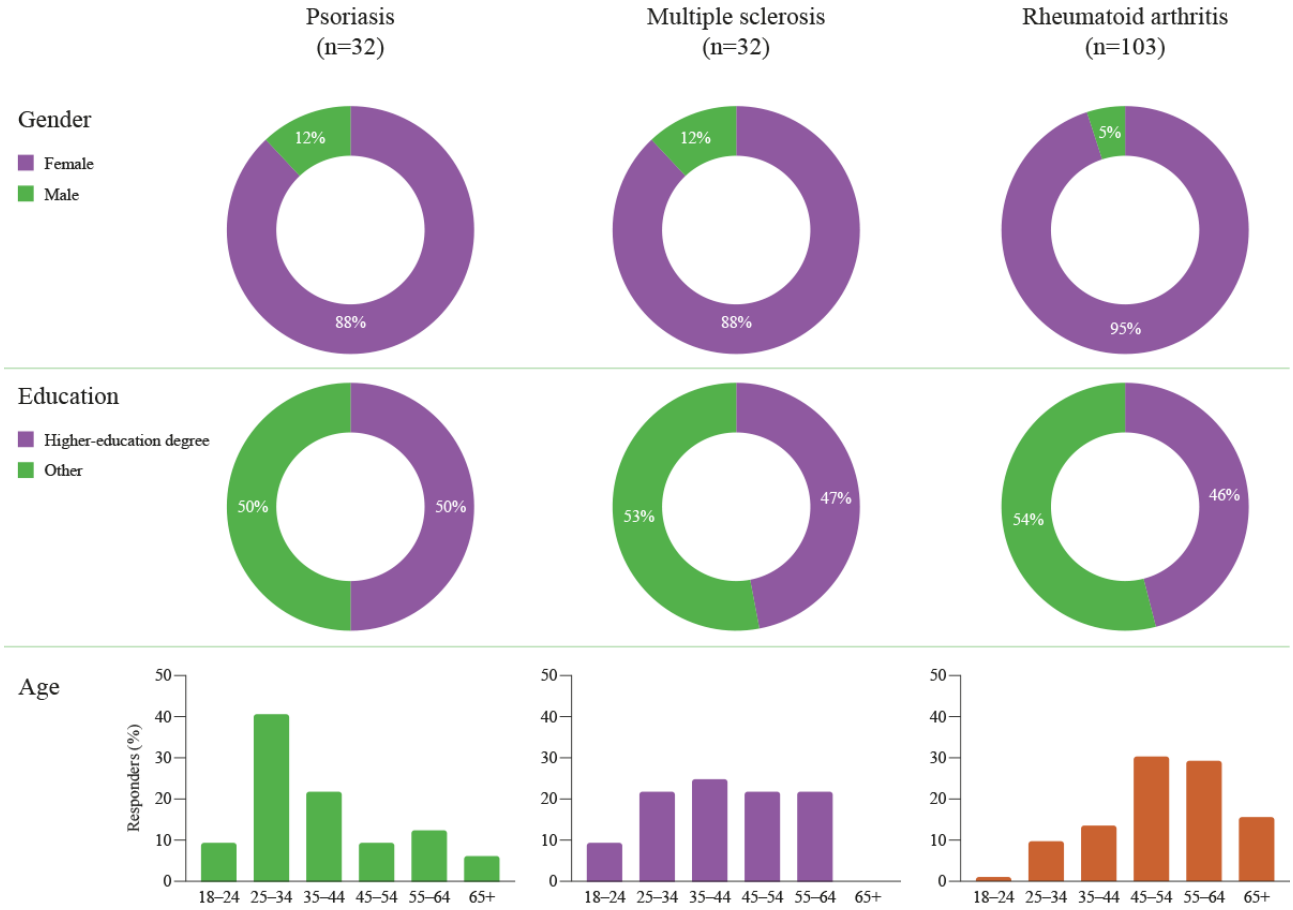
This article was prepared in accordance with the Checklist for Reporting Results of Internet E-Surveys (CHERRIES; see Multimedia Appendix 1) [33].

Results

Survey Respondents

In total, 167 survey responses were received for the 3 surveys (psoriasis, n=32; MS, n=32; RA, n=103; Figure 2). The survey completion rates were 84% (27/32) for psoriasis, 81% (26/32) for MS, and 90% (93/103) for RA. Those who responded to the survey were mainly women (psoriasis, 28/32, 88%; MS, 28/32, 88%; RA, 97/102, 95%), and approximately half were educated to university (higher-education) degree level (psoriasis, 16/32, 50%; MS, 15/32, and 47%; RA, 47/102, 46%). Age ranges for respondents were as expected for each of the 3 disease states.

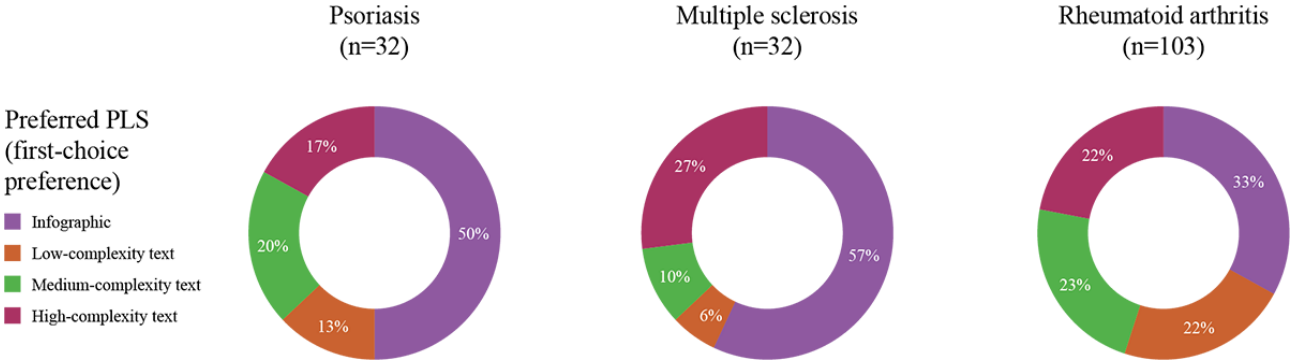
Figure 2. Demographics of survey respondents across different chronic disease states.



Primary Analysis

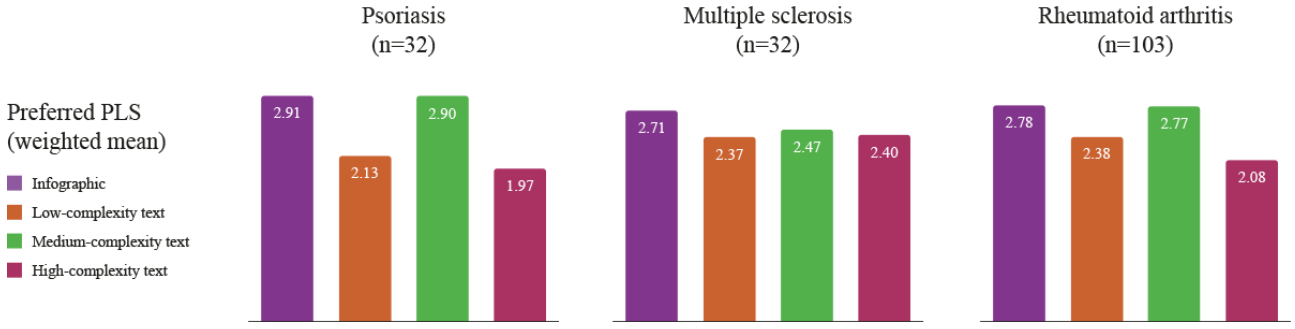
Across all 3 disease states, the infographic was the first-choice PLS format for most respondents (psoriasis, 15/30, 50%; MS, 17/30, 57%; RA, 33/100, 33%), whereas the low-complexity text-only PLS was the least preferred first-choice format (psoriasis, 4/30, 13%; MS, 2/30, 6%; RA, 22/100, 22%; see Figure 3).

Figure 3. Respondents' first-choice preference of PLS format for the different disease states. PLS: plain language summary.



Similarly, results from the weighted-average preference score data demonstrated that the infographic (psoriasis 2.91; MS 2.71; RA 2.78) and medium-complexity text-only PLSs (reading age 14-17 years, US Grade 9-11; psoriasis 2.90; MS 2.47; RA 2.77) were the 2 most popular PLS formats across all 3 diseases analyzed (Figure 4).

Figure 4. Weighted mean scores for preferred PLS format. PLS: plain language summary.



Subgroup Analysis

Among those who chose the infographic format as their first-choice preference, the majority for both the psoriasis and MS groups were in the younger age category of 18 to 34 years (psoriasis 10/15, 67% and MS 8/17, 47%). In contrast, for the RA group, the preference for the infographic format was similar among middle-aged (35-54 years; 16/33, 49%) and older respondents (≥55 years; 14/33, 42%; see Multimedia Appendix 8). When analyzed by education level, we found that about half the respondents who preferred the infographic format for psoriasis and MS had a higher-education degree (psoriasis 8/15, 53% and MS 8/17, 47%); this proportion was lower for RA (12/33, 36%; see Multimedia Appendix 8).

Of the respondents who preferred the medium-complexity text-only PLS format, the majority in each group were ≥55 years of age for MS (2/4, 50%) and RA (14/23, 61%); however, for psoriasis, the preference was equal across all 3 age groups analyzed (2/6, 33%; see Multimedia Appendix 8). The proportion of respondents with a higher-education degree and those who preferred the medium-complexity text-only PLS was similar for psoriasis (3/6, 50%) and MS (2/4, 50%), but higher for RA (14/23, 61%).

Free-text Feedback Samples

As part of the survey, respondents were able to provide free-text feedback regarding the different PLS formats (Multimedia Appendix 9). Responses were almost the same across the 3 disease types. In general, infographic and medium-complexity text-only PLS formats were praised for their clear and concise presentation, while maintaining the relevant level of information. In contrast, the high-complexity and low-complexity text-only PLS formats were criticized for the use of jargon or oversimplification, respectively.

Reasons (verbatim) provided by the survey respondents for specific preference for the infographic or medium-complexity text-only PLS formats are listed below:

> *…[infographic] helpful in getting statistical information across.*
>
> *...[infographic] clear, concise, and easy to understand.*
>
> *…graphic summary was accessible almost at a glance.*
>
> *…[infographic] well detailed and a lot easier to read than loads of text.*
>
> *[medium-complexity] summary combined straightforward language with enough information.*
>
> *[medium-complexity] text was well detailed and in a language that was easy to understand by anyone.*
>
> *[medium-complexity] would 'fill-in-the-gaps' and provide the detail that the [infographic] summary by its very nature could not.*

Reasons (verbatim) provided by the survey respondents that suggest the high-complexity text-only and low-complexity text-only PLS formats were less popular are listed below:

> *The [high-complexity text] was way too hard to understand what they were saying, too many big words.*
>
> *[High-complexity] summary was quite complex to understand.*
>
> *[High-complexity] requires greater concentration and previous experience of medical terminology, e.g., AE – not all would know that this means adverse event.*
>
> *I ranked the [low-complexity] text summary 4th because, although it was short and easy to read, it did not give me the pertinent statistical data from which the conclusion was drawn.*
>
> *I thought the [low-complexity] text summary had been simplified to the extent that it lost some meaning.*
>
> *I personally hate dumbed-down items… Difficult subjects should be explained…*

Discussion

Our findings showed a clear preference for an infographic PLS format among the 3 disease states we assessed (psoriasis, MS, and RA). Medium-complexity (reading age 14-17 years, US Grade 9-11) was the most preferred text-based format. The main reasons cited for preferring these formats were that the information presented was clear, concise, easy to understand, and included relevant detail, without oversimplification of the content. The majority of respondents were women, and approximately half had a university-level degree. Preferences remained the same regardless of education status; however, younger respondents were more likely to prefer the infographic to the text-based format.

PLSs are an important tool for improving health literacy. Research has indicated that most nonexperts express difficulty

understanding medical and scientific texts, especially when reading text describing complex clinical research [34].

In recent years, there have been initiatives to improve the reach of information meant for a nonexpert audience. In 2010, the United States Congress recognized the need for the use of plain language when communicating information intended for the public [35]. For example, The National Action Plan to Improve Health Literacy aims to engage organizations, professionals, policy makers, communities, individuals, and families in a linked, multi-sector effort to improve the understanding of basic health information [36]. The plan is based on two "core principles": (1) all people have the right to health information that helps them make informed decisions, and (2) health information should be delivered in ways that are easy to understand and that improve health, longevity, and quality of life. Moreover, the US National Institutes of Health aim to broaden the reach of health information to all Americans by communicating research results in terms that are easy to understand [7]. Similar initiatives are taking place in the European Union, where CTS must accompany clinical trial results for laypersons [15]. CTSs aim to be accessible to the general public as young as 12 years [15]. Health literacy is important to young patients; therefore, supporting them in understanding health issues can empower them to take control of their health and provide the information they need to seek appropriate services [37].

Additionally, in 2017, to increase public access to healthcare information, the nonprofit organization eLife compiled a list of organizations that provide PLSs of published scientific research—including more than 50 medical or scientific journals [38]. We now estimate that more than 250 medical journals now facilitate the provision of PLSs; however, wide variations still exist in the terms used for defining a PLS, the presentation format, the platforms where they are located (eg, journal website, Figshare, or Kudos), and how the reader may discover them (eg, via a PubMed search) [39].

To standardize PLS formats, readability scores and formulas (within applications like Microsoft Word or by using web-based tools [24]) have been used to assess the complexity of text; however, such a metrics-based approach fails to incorporate individual preferences regarding information delivery and overlooks the importance of engaging the audience or assessing whether the information will be interpreted as intended. Furthermore, although PLS formats are far easier to read than other traditional formats, the level of literacy preferred by the population surveyed in our study was relatively higher than the recommended reading age for a CTS and health-related information (generally 10-12 years of age, US Grade 5-7) [4,15,40,41].

The best way of presenting research results to different audiences remains unclear. Few studies have investigated this topic, which indicates that scientific findings can be difficult to interpret [9]. Furthermore, other research has identified that, in addition to PLS, video abstracts may be preferable than published text or graphical abstracts [42]. Our study provides valuable insight and direction for how PLSs may be formatted and presented to communicate original medicine-based research to a broader audience. Through the use of a survey, people who responded were able to state their preferences for PLS formats and also provide reasons for their preference, referring to the key factors that dictate how they wish to receive scientific information. The participants' preference for both the infographic and medium-complexity text-based PLS format was based on clarity and concise distribution of information, without sacrificing key details. The high-complexity text-based PLS was thought to have excessive use of jargon, requiring a scientific background to appreciate the information adequately. Conversely, participants were dissatisfied with the low-complexity PLS format, considering it too simple and not having enough substance.

Limitations

The overall sample may have been a more educated population than the general public, being sourced from patient groups and those who regularly use the internet and social media. We consider this more educated population to be representative of the technology-competent, information-seeking individuals most likely to be sourcing and reading PLSs; however, it does not necessarily capture the preferences of audiences who are less technologically aware and who may still benefit from exposure to clinical research through reading PLSs.

There was a notable gender imbalance within all 3 subpopulations surveyed, with approximately 90% of the respondents being women in each case. Women are more likely to experience RA and MS than men, whereas the prevalence is about equal for those with psoriasis [43-45]; however, the imbalance was far more significant in this survey than that observed in the real-world setting, which we are unable to explain. The survey also did not capture whether those who responded were patients or caregivers, which could have provided more context to the results.

Although the samples for each disease included in the survey provided enough data to generate meaningful results, the number of people who responded to the psoriasis and MS surveys were only a third of those who responded to the RA survey. Results of the subgroup analyses should be treated with caution due to the low number of respondents in each subgroup, particularly from the psoriasis and MS populations.

Since developing the PLS for each of the 3 source articles surveyed, our understanding and application of best practices in plain language writing for publications have advanced. If we were to repeat this project, we would apply more principles outlined in the tools to help guide plain language writing [6,14,18-20] to the development of PLSs used in the analyses.

Furthermore, although we used a web-based tool to assess readability [24], we recommend caution when using such tools for content that is confidential and where the security of the tool has not been verified.

Conclusions

Audience preferences should be accounted for when preparing a PLS to supplement an original peer-reviewed research article. However, oversimplification of text can be viewed negatively, and infographic versions or medium-complexity text appear to

be the most popular. Further research would be useful to expand both the scope of the therapy areas covered and the profile of those surveyed to include other nonexpert populations and healthcare professionals from other fields of study. It would also be of interest to evaluate the understanding of the information presented in a PLS rather than focus on the preferred format alone. Training at professional societies such as the International Society for Medical Publication Professionals [46] and the widespread use of additional tools now available to guide the effective production of PLSs [6,14,19-21] will help facilitate this.

Acknowledgments

This research was sponsored by the Manchester Metropolitan University and was supported by CMC Connect, McCann Health Medical Communications. The authors thank all individuals who participated in the survey. Medical writing support, under the direction of the authors, was provided by Rachel Janes and Dominic Singson of CMC Connect, McCann Health Medical Communications, in accordance with Good Publication Practice guidelines.

Authors' Contributions

All authors were involved in defining and conducting the research and preparing the manuscript.

Conflicts of Interest

JG and CS are employees of CMC Connect, McCann Health Medical Communications, and support clients in the development of plain language summaries of publications.

Multimedia Appendix 1
Plain language summary of this article in an infographic format.
[PDF File (Adobe PDF File), 227 KB - jmir_v24i1e22122_app1.pdf]

Multimedia Appendix 2
High-, medium-, and low-complexity text-based plain language summaries for psoriasis.
[PDF File (Adobe PDF File), 899 KB - jmir_v24i1e22122_app2.pdf]

Multimedia Appendix 3
High-, medium-, and low-complexity text-based plain language summaries for multiple sclerosis.
[PDF File (Adobe PDF File), 516 KB - jmir_v24i1e22122_app3.pdf]

Multimedia Appendix 4
High-, medium-, and low-complexity text-based plain language summaries for rheumatoid arthritis.
[PDF File (Adobe PDF File), 284 KB - jmir_v24i1e22122_app4.pdf]

Multimedia Appendix 5
Summary of readability scores and other features of text-based plain language summaries for all 3 sample articles evaluated. PLS: plain language summary.
[PDF File (Adobe PDF File), 380 KB - jmir_v24i1e22122_app5.pdf]

Multimedia Appendix 6
Survey questionnaire for all 3 disease states evaluated.
[PDF File (Adobe PDF File), 100 KB - jmir_v24i1e22122_app6.pdf]

Multimedia Appendix 7
UK-based patient associations and Facebook groups used to identify survey respondents.
[PDF File (Adobe PDF File), 175 KB - jmir_v24i1e22122_app7.pdf]

Multimedia Appendix 8
Subgroup analyses. MS: multiple sclerosis; PLS: plain language summary; RA: rheumatoid arthritis.
[PDF File (Adobe PDF File), 14377 KB - jmir_v24i1e22122_app8.pdf]

Multimedia Appendix 9

Free-text feedback on reasons for the preferred PLS format. PLS: plain language summary.
[PDF File (Adobe PDF File), 364 KB - jmir_v24i1e22122_app9.pdf]

Multimedia Appendix 10
Checklist for Reporting Results of Internet E-Surveys (CHERRIES).
[PDF File (Adobe PDF File), 163 KB - jmir_v24i1e22122_app10.pdf]

References

1. Barnes A, Patrick S. Lay summaries of clinical study results: an overview. Pharmaceut Med 2019 Aug;33(4):261-268. [doi: 10.1007/s40290-019-00285-0] [Medline: 31933186]
2. Stossel LM, Segar N, Gliatto P, Fallar R, Karani R. Readability of patient education materials available at the point of care. J Gen Intern Med 2012;27(9):1165-1170 [FREE Full text] [doi: 10.1007/s11606-012-2046-0] [Medline: 22528620]
3. Koh HK, Brach C, Harris LM, Parchman ML. A proposed 'health literate care model' would constitute a systems approach to improving patients' engagement in care. Health Aff (Millwood) 2013;32(2):357-367 [FREE Full text] [doi: 10.1377/hlthaff.2012.1205] [Medline: 23381529]
4. Karačić J, Dondio P, Buljan I, Hren D, Marušić A. Languages for different health information readers: multitrait-multimethod content analysis of Cochrane systematic reviews textual summary formats. BMC Med Res Methodol 2019 Apr 05;19(1):75 [FREE Full text] [doi: 10.1186/s12874-019-0716-x] [Medline: 30953453]
5. State of Science Index Survey. 3M. URL: https://www.3m.com/3M/en_US/state-of-science-index-survey/2019-summary/ [accessed 2019-05-27]
6. Duke M. How to write a lay summary. Digital Curation Centre. 2012. URL: https://www.dcc.ac.uk/sites/default/files/documents/publications/HowToLaySummariesDec2012.pdf [accessed 2021-05-27]
7. Clear communication. National Institutes of Health – Office of Communications and Public Liaison. 2020. URL: https://www.nih.gov/institutes-nih/nih-office-director/office-communications-public-liaison/clear-communication [accessed 2021-05-27]
8. Kerwer M, Chasiotis A, Stricker J, Günther A, Rosman T. Straight from the scientist's mouth—plain language summaries promote laypeople's comprehension and knowledge acquisition when reading about individual research findings in psychology. Psychology 2021;7(1):18898 [FREE Full text] [doi: 10.1525/collabra.18898]
9. Buljan I, Malički M, Wager E, Puljak L, Hren D, Kellie F, et al. No difference in knowledge obtained from infographic or plain language summary of a Cochrane systematic review: three randomized controlled trials. J Clin Epidemiol 2018;97:86-94 [FREE Full text] [doi: 10.1016/j.jclinepi.2017.12.003] [Medline: 29269021]
10. Pushparajah DS, Manning E, Michels E, Arnaudeau-Bégard C. Value of developing plain language summaries of scientific and clinical articles: a survey of patients and physicians. Ther Innov Regul Sci 2018;52(4):474-481. [doi: 10.1177/2168479017738723] [Medline: 29714545]
11. Salita JT. Writing for lay audiences: a challenge for scientists. Medical Writing 2015;24(4):183-189. [doi: 10.1179/2047480615z.000000000320]
12. Blyth J, Gaskarth M, Plant A, Woods K. We need to talk about PLS...exploring the opportunities of plain-language summaries. ISMPP Newsletter. 2019. URL: https://ismpp-newsletter.com/2019/05/01/we-need-to-talk-about-plsexploring-the-opportunities-of-plain-language-summaries/ [accessed 2021-05-27]
13. Warde F, Papadakos J, Papadakos T, Rodin D, Salhia M, Giuliani M. Plain language communication as a priority competency for medical professionals in a globalized world. Can Med Educ J 2018;9(2):e52-e59 [FREE Full text] [Medline: 30018684]
14. Good Lay Summary Practice. Clinical Trials Expert Group. 2021 Oct 04. URL: https://ec.europa.eu/health/sites/default/files/files/eudralex/vol-10/glsp_en.pdf [accessed 2021-11-26]
15. Summaries of clinical trials for laypersons. European Commission. 2017. URL: https://ec.europa.eu/health/sites/health/files/files/eudralex/vol-10/2017_01_26_summaries_of_ct_results_for_laypersons.pdf [accessed 2021-05-27]
16. Plain language summary guidance document submitted to the FDA–open for comments. Multi-regional Clinical Trials Center. 2017. URL: https://mrctcenter.org/blog/news/plain-language-summary-guidance-document-submitted-fda-open-comments/ [accessed 2021-05-27]
17. Plain English summaries: What is a plain English summary? National Institute for Health Research. 2021. URL: https://www.invo.org.uk/resource-centre/plain-english-summaries/what-is-a-plain-english-summary/ [accessed 2021-09-09]
18. Lobban D, Gardner J, Matheis R. Plain language summaries of publications of company-sponsored medical research: what key questions do we need to address? Curr Med Res Opin. Epub ahead of print posted online on November 4, 2021 :1-12 [FREE Full text] [doi: 10.1080/03007995.2021.1997221] [Medline: 34736362]
19. Plain language summaries (PLS) of publications toolkit: a best-practice resource for PLS of peer-reviewed publications and congress abstracts. Envision Pharma Group. 2018. URL: https://www.envisionthepatient.com/plstoolkit/ [accessed 2021-05-27]

20. Methodological Expectations of Cochrane Intervention Reviews (MECIR): Standards for the reporting of plain language summaries in new Cochrane Intervention Reviews. Cochrane Methods. 2013. URL: https://methods.cochrane.org/sites/default/files/public/uploads/pleacs_2019.pdf [accessed 2021-09-09]
21. Rosenberg A, Baróniková S, Feighery L, Gattrell W, Olsen RE, Watson A, et al. Open Pharma recommendations for plain language summaries of peer-reviewed medical journal publications. Curr Med Res Opin 2021;37(11):2015-2016. [doi: 10.1080/03007995.2021.1971185] [Medline: 34511020]
22. Chamberlain James L, Bharadia T. Lay summaries and writing for patients: Where are we now and where are we going? Medical Writing 2019;28(3):46-51.
23. Crick K, Hartling L. Preferences of knowledge users for two formats of summarizing results from systematic reviews: infographics and critical appraisals. PLoS One 2015;10(10):e0140029 [FREE Full text] [doi: 10.1371/journal.pone.0140029] [Medline: 26466099]
24. Readability formulas and the active role of the reader. Readability Formulas. 2019. URL: http://www.readabilityformulas.com/articles/readability-formulas-and-the-active-role-of-the-reader.php [accessed 2021-05-27]
25. Flesch R. A new readability yardstick. J Appl Psychol 1948;32(3):221-233. [Medline: 18867058]
26. Kincaid JP, Fishburne R, Rogers R, Chissom B. Derivation of new readability formulas (automated readability index, fog count, and Flesch reading ease formula) for Navy enlisted personnel. Springfield, Virginia: National Technical Information Service; 1975:1-75.
27. The 2011 Skills for Life Survey: A Survey of Literacy, Numeracy and ICT Levels in England. Department for Business Innovation and Skills. London: Department for Business Innovation and Skills; 2011. URL: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/36000/12-p168-2011-skills-for-life-survey.pdf [accessed 2021-09-09]
28. Content design: planning, writing and managing content. UK Government Digital Service. 2016 Feb 25. URL: https://www.gov.uk/guidance/content-design [accessed 2021-09-09]
29. Warren RB, Mrowietz U, von Kiedrowski R, Niesmann J, Wilsmann-Theis D, Ghoreschi K, et al. An intensified dosing schedule of subcutaneous methotrexate in patients with moderate to severe plaque-type psoriasis (METOP): a 52 week, multicentre, randomised, double-blind, placebo-controlled, phase 3 trial. Lancet 2017 Feb 04;389(10068):528-537. [doi: 10.1016/S0140-6736(16)32127-4] [Medline: 28012564]
30. Hauser SL, Bar-Or A, Comi G, Giovannoni G, Hartung H, Hemmer B, OPERA IOPERA II Clinical Investigators. Ocrelizumab versus interferon beta-1a in relapsing multiple sclerosis. N Engl J Med 2017 Jan 19;376(3):221-234. [doi: 10.1056/NEJMoa1601277] [Medline: 28002679]
31. Aletaha D, Bingham CO, Tanaka Y, Agarwal P, Kurrasch R, Tak PP, et al. Efficacy and safety of sirukumab in patients with active rheumatoid arthritis refractory to anti-TNF therapy (SIRROUND-T): a randomised, double-blind, placebo-controlled, parallel-group, multinational, phase 3 study. Lancet 2017 Mar 25;389(10075):1206-1217. [doi: 10.1016/S0140-6736(17)30401-4] [Medline: 28215362]
32. Survey Monkey. About the cookies we use.: Survey Monkey URL: https://help.surveymonkey.com/articles/en_US/kb/About-the-cookies-we-use [accessed 2021-09-09]
33. Eysenbach G. Improving the quality of web surveys: the Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res 2004 Sep 29;6(3):e34 [FREE Full text] [doi: 10.2196/jmir.6.3.e34] [Medline: 15471760]
34. Smith CA, Hetzel S, Dalrymple P, Keselman A. Beyond readability: investigating coherence of clinical text for consumers. J Med Internet Res 2011 Dec 02;13(4):e104 [FREE Full text] [doi: 10.2196/jmir.1842] [Medline: 22138127]
35. United States Congress. Plain Writing Act of 2010. Public Law 111-274-Oct. 13 2010. 124 STAT. 2861. Congressional record, Vol. 156 (2010). 2010 Oct 13. URL: https://www.govinfo.gov/content/pkg/PLAW-111publ274/pdf/PLAW-111publ274.pdf [accessed 2021-11-26]
36. National action plan to improve health literacy. US Department of Health and Human Services, Office of Disease Prevention and Health Promotion. Washington, DC; 2010. URL: https://health.gov/communication/initiatives/health-literacy-action-plan.asp [accessed 2021-05-27]
37. Hagell A. Association for Young People's Health. Promoting young people's health literacy and understanding their help-seeking behaviour. URL: http://www.youngpeopleshealth.org.uk/wp-content/uploads/2015/07/628_Health-literacy-EE-format-Updated-by-AH-march.pdf [accessed 2021-05-27]
38. Plain-language summaries: journals and other organizations that produce plain-language summaries. eLIFE. 2017. URL: https://elifesciences.org/inside-elife/5ebd9a3f/plain-language-summaries-journals-and-other-organizations-that-produce-plain-language-summaries [accessed 2021-05-27]
39. FitzGibbon H, King K, Piano C, Wilk C, Gaskarth M. Where are biomedical research plain-language summaries? Health Sci Rep 2020 Sep;3(3):e175 [FREE Full text] [doi: 10.1002/hsr2.175] [Medline: 32789193]
40. Rowlands G, Bond B, Shay M. The Information Standard Workshop: Understanding Health Literacy – best practice in developing and testing health and care information. The Information Standard. London: NHS England; 2013 Oct 10. URL: https://www.england.nhs.uk/wp-content/uploads/2017/02/tis-standard-health-literacy-wrkshp-handout-101013.pdf [accessed 2021-05-27]

41. Badarudeen S, Sabharwal S. Assessing readability of patient education materials: current role in orthopaedics. Clin Orthop Relat Res 2010 Oct;468(10):2572-2580 [FREE Full text] [doi: 10.1007/s11999-010-1380-y] [Medline: 20496023]
42. Bredbenner K, Simon SM. Video abstracts and plain language summaries are more effective than graphical abstracts and published abstracts. PLoS One 2019;14(11):e0224697 [FREE Full text] [doi: 10.1371/journal.pone.0224697] [Medline: 31743342]
43. Humphreys JH, Verstappen SMM, Hyrich KL, Chipping JR, Marshall T, Symmons DPM. The incidence of rheumatoid arthritis in the UK: comparisons using the 2010 ACR/EULAR classification criteria and the 1987 ACR classification criteria. Results from the Norfolk Arthritis Register. Ann Rheum Dis 2013 Aug;72(8):1315-1320 [FREE Full text] [doi: 10.1136/annrheumdis-2012-201960] [Medline: 22945499]
44. Mackenzie IS, Morant SV, Bloomfield GA, MacDonald TM, O'Riordan J. Incidence and prevalence of multiple sclerosis in the UK 1990-2010: a descriptive study in the General Practice Research Database. J Neurol Neurosurg Psychiatry 2014 Jan;85(1):76-84 [FREE Full text] [doi: 10.1136/jnnp-2013-305450] [Medline: 24052635]
45. Springate DA, Parisi R, Kontopantelis E, Reeves D, Griffiths CEM, Ashcroft DM. Incidence, prevalence and mortality of patients with psoriasis: a U.K. population-based cohort study. Br J Dermatol 2017 Mar;176(3):650-658 [FREE Full text] [doi: 10.1111/bjd.15021] [Medline: 27579733]
46. The International Society for Medical Publication Professionals. URL: http://www.ismpp.org/ [accessed 2021-09-09]

Abbreviations

CHERRIES: Checklist for Reporting Results of Internet E-Surveys
CTS: clinical trial summary
GCSE: General Certificate of Secondary Education
GLSP: Good Lay Summary Practice
MS: multiple sclerosis
PLS: plain language summary
RA: rheumatoid arthritis

Edited by R Kukafka, G Eysenbach; submitted 23.07.20; peer-reviewed by R Krukowski, J Li; comments to author 02.08.20; revised version received 16.10.20; accepted 29.09.21; published 11.01.22.

Please cite as:
Martínez Silvagnoli L, Shepherd C, Pritchett J, Gardner J
Optimizing Readability and Format of Plain Language Summaries for Medical Research Articles: Cross-sectional Survey Study
J Med Internet Res 2022;24(1):e22122
URL: https://www.jmir.org/2022/1/e22122
doi:10.2196/22122
PMID:35014966

Original Paper

Validity Testing and Cultural Adaptation of the eHealth Literacy Questionnaire (eHLQ) Among People With Chronic Diseases in Taiwan: Mixed Methods Study

Yu-Chi Chen[1], PhD; Christina Cheng[2], PhD; Richard H Osborne[2,3], PhD; Lars Kayser[3], MD, PhD; Chieh-Yu Liu[4], PhD; Li-Chun Chang[5], PhD

[1]Institute of Clinical Nursing, College of Nursing, National Yang Ming Chiao Tung University, Taipei, Taiwan

[2]Centre for Global Health and Equity, Swinburne University of Technology, Melbourne, Australia

[3]Department of Public Health, University of Copenhagen, Copenhagen, Denmark

[4]Department of Health Care Management, National Taipei University of Nursing and Health Sciences, Taipei, Taiwan

[5]School of Nursing, Chang Gung University of Science and Technology, Tao-Yuan, Taiwan

Corresponding Author:
Yu-Chi Chen, PhD
Institute of Clinical Nursing
College of Nursing
National Yang Ming Chiao Tung University
No.155, Section 2, Li-Nong Street
Beitou District
Taipei, 112
Taiwan
Phone: 886 228267093
Fax: 886 228202487
Email: ycchen2@nycu.edu.tw

Abstract

Background: Advancements in digital technologies seek to promote health and access to services. However, people lacking abilities and confidence to use technology are likely to be left behind, leading to health disparities. In providing digital health services, health care providers need to be aware of users' diverse electronic health (eHealth) literacy to address their particular needs and ensure equitable uptake and use of digital services. To understand such needs, an instrument that captures users' knowledge, skills, trust, motivation, and experiences in relation to technology is required. The eHealth Literacy Questionnaire (eHLQ) is a multidimensional tool with 7 scales covering diverse dimensions of eHealth literacy. The tool was simultaneously developed in English and Danish using a grounded and validity-driven approach and has been shown to have strong psychometric properties.

Objective: This study aims to translate and culturally adapt the eHLQ for application among Mandarin-speaking people with chronic diseases in Taiwan and then undertake a rigorous set of validity-testing procedures.

Methods: The cross-cultural adaptation of the eHLQ included translation and evaluation of the translations. The measurement properties were assessed using classical test theory and item response theory (IRT) approaches. Content validity, known-group validity, and internal consistency were explored, as well as item characteristic curves (ICCs), item discrimination, and item location/difficulty.

Results: The adapted version was reviewed, and a recommended forward translation was confirmed through consensus. The tool exhibited good content validity. A total of 420 people with 1 or more chronic diseases participated in a validity-testing survey. The eHLQ showed good internal consistency (Cronbach α=.75-.95). For known-group validity, all 7 eHLQ scales showed strong associations with education. Unidimensionality and local independence assumptions were met except for scale 2. IRT analysis showed that all items demonstrated good discrimination (range 0.27-12.15) and a good range of difficulty (range 0.59-1.67) except for 2 items in scale 7.

Conclusions: Using a rigorous process, the eHLQ was translated from English into a culturally appropriate tool for use in the Mandarin language. Validity testing provided evidence of satisfactory-to-strong psychometric properties of the eHLQ. The 7

scales are likely to be useful research tools for evaluating digital health interventions and for informing the development of health technology products and interventions that equitably suit diverse users' needs.

(J Med Internet Res 2022;24(1):e32855) doi:10.2196/32855

KEYWORDS

chronic illness; eHealth literacy questionnaire; eHLQ; validation; cultural adaptation; eHealth

Introduction

In societies with a rapid ongoing service transformation of health care to be more digitally supported and expectations of higher community involvement, it is necessary that people be actively supported to participate in their own care, including engagement with electronic health (eHealth) care resources. People need to be able to obtain relevant health information and support from web-based services, use technology for health management, and receive appropriate care from eHealth service systems [1-3]. In this era of eHealth care management, understanding the eHealth literacy (eHL) of service users is important to ensure they can equitably benefit from and take advantage of the digital services and health technologies [4,5]. People with a range of eHL skills are more likely to engage in eHealth resources, leading to improved knowledge, skills, and confidence to actively manage their health condition [6-9]. Conversely, people with low eHL may not be able to understand, access, and use health and care services and health information, leading to suboptimal disease self-management, increased vulnerability, and poor health outcomes [3,6,10].

Digital health care service solutions need to recognize and respond to users' personal goals, values, and competence in a sociotechnical context. In terms of the solution's function and interface design, the solution must not only serve the goals of health service providers but also satisfy diverse users' needs across their eHL levels; this will increase the benefits and upscale the benefits of the solution [11-13].

The concept of eHL was introduced in the Web 1.0 era by Norman and Skinner [1,14] in 2006 and was described as "the ability to seek, find, understand, and appraise health information from electronic sources and apply the knowledge gained to addressing or solving a health problem." Over the past 2 decades, health services have become more complex and interactive with the expectation that users be active in managing their own condition using digital services. As such, new and more comprehensive tools to measure eHL are required [15]. In response to the advances in health technology, Norgaard et al [5] proposed in 2015 the eHealth Literacy Framework (eHLF), which comprises 7 dimensions of eHL.

The eHLF, developed using a grounded validity-driven approach [16] with extensive international consultation with service users, health professionals, researchers, and technology experts, provides a contemporary and comprehensive map of an individual's technology health literacy. The eHLF covers knowledge and skills, the eHealth system's attributes, and how an individual interacts with the system. Subsequently, the eHealth Literacy Questionnaire (eHLQ) was developed based on the eHLF in Danish and English. The tool was tested in Denmark in a large sample of people with chronic diseases and the general population. The questionnaire was found to have a wide range of excellent psychometric properties [17,18].

Today, the far-reaching nature of the digital environment with the internet and cloud technology makes services and information borderless, and issues associated with eHL can also have global ramifications. What is the relationship between eHL and health care behaviors? What is the difference in eHL levels between the people of Taiwan or China and those of other countries? To explore these issues, an appropriate and psychometrically sound evaluation tool is required. Although the eHLQ has undergone validity testing in Denmark in both health and community settings and is available in Danish and English, a Chinese version for use in Taiwan is required. Given that Denmark and Taiwan have different health care systems and that items and constructs may be subject to differential cultural and linguistic interpretations, it is important that careful translation and cultural adaption, as well as psychometric testing, be undertaken to inform researchers, clinicians, and health system managers in Taiwan and other Mandarin-speaking areas. The aim of this study was to translate and culturally adapt the eHLQ from English to Chinese and evaluate its cultural and psychometric properties in a group of Mandarin-speaking people with chronic diseases.

Methods

Study Design

This was a 2-phase study. Phase 1 involved the translation and cultural adaptation of the eHLQ for application in the Mandarin language. In phase 2, the Chinese version was psychometrically tested among people with chronic diseases using classical test theory and item response theory (IRT) approaches. IRT, also known as latent response theory, refers to a family of mathematical models that seek to explain the relationship between latent traits (unobservable characteristic or attribute) and their manifestations.

eHealth Literacy Questionnaire

The eHLQ has 35 items representing 7 scales that cover the eHLF dimensions: (1) using technology to process health information, (2) understanding of health concepts and language, (3) ability to actively engage with digital services, (4) feel safe and in control, (5) motivated to engage with digital services, (6) access to digital services that work, and (7) digital services that suit individual needs [5,17]. The scale names and construct definitions [17] are shown in Multimedia Appendix 1. Each scale has 4-6 items with 4-point response options: strongly disagree, disagree, agree, and strongly agree, with an assigned value of 1-4, respectively. Scale scores are calculated by

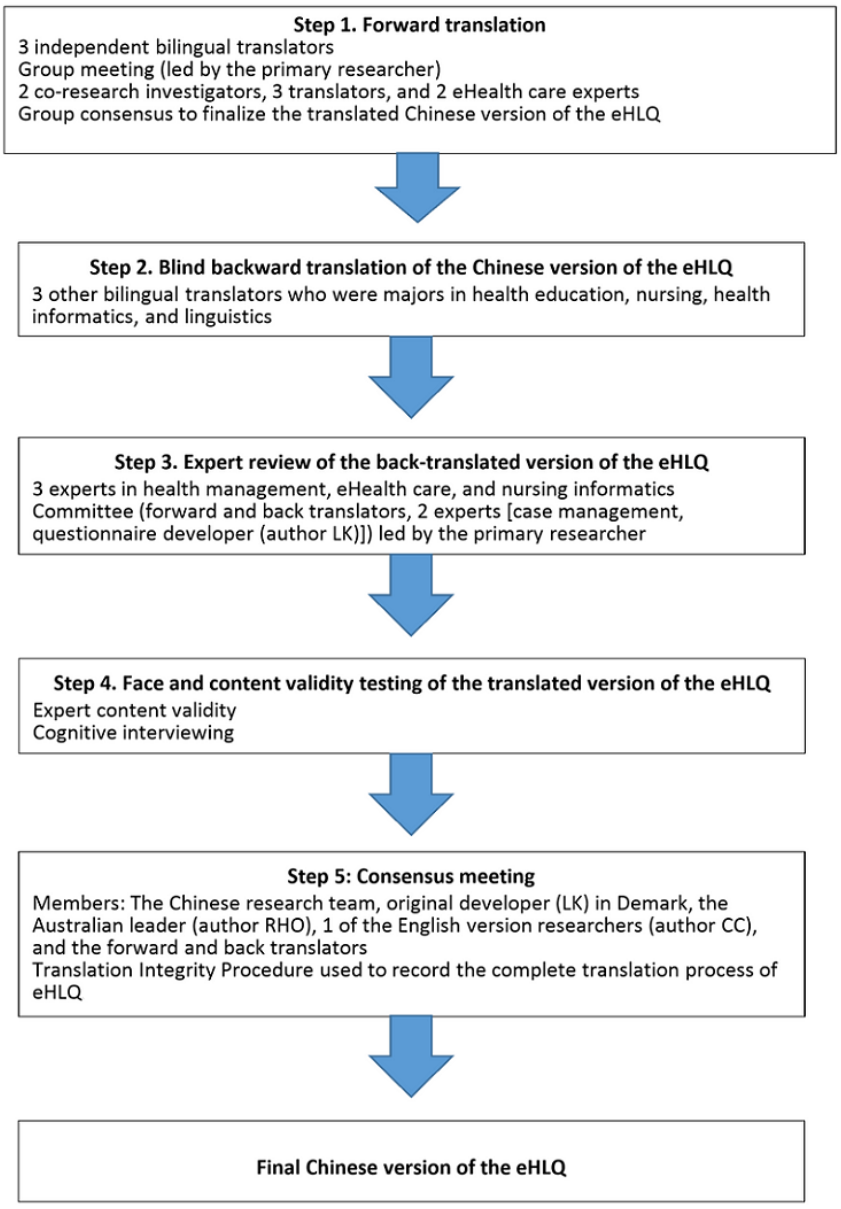
averaging the items scores within each scale with equal weighting, generating scale scores that range from 1 to 4.

Phase 1. Translation and Cultural Adaptation

Initial Translation Process

The translation of a questionnaire should not only include a textual change but also consider cultural equivalence and applicability [19-22]. Therefore, this study was designed according to the guidelines for instrument translation, adaptation, and validation proposed by Sousa and Rojjanasrirat [21] and Hall et al [19] and, in particular, the Translation Integrity Procedure developed by Hawkins and Osborne [23] for the eHLQ. The Translation Integrity Procedure ensures that the language translation and cultural adaptation follow detailed item intent descriptions and seek to ensure cultural suitability and measurement equivalence (ie, whether each concept in the translated version is the same strength as that in the original version) [23]. See Figure 1.

Figure 1. The translation process. eHLQ: eHealth Literacy Questionnaire.



Three bilingual translators in both English and Mandarin independently translated the English version of the eHLQ into traditional Chinese, including analysis of cultural appropriateness and measurement equivalence by 5 experts in health care and informatics. Words and phrases that diverged from the original intent were discussed with the translators. Finally, minor modifications of the translated version were incorporated to preserve semantic and idiomatic equivalence in traditional Chinese characters for Mandarin speakers. The translated version was then translated back into English by 3 translators with linguistic qualifications or health or technology qualifications. Two were native English speakers. The forward translation was discussed until a linguistically and culturally equivalent meaning was achieved between the source (guided

by item intent and overall construct meaning) and the forward Chinese version.

Content Validity and Cognitive Interviewing

Nine experts in nursing, medical practice, public health, informatics, eHealth care, and patient health education were invited to evaluate the content validity of the translated eHLQ. Based on the operational definitions and the item intent of the eHLQ (refer to Kayser et al [17] and the Translation Integrity Procedure [23]), experts evaluated whether the items were representative of the construct and were clearly stated. The level of representativeness was rated on a scale of 1-4, with 4 being representative, 3 being representative but needing minor revision, 2 being representative but needing major revision, and 1 not being representative. Furthermore, the level of clarity was rated on a similar scale, with 4 being clear, 3 being clear but needs minor revision, 2 being clear but needing major revision, and 1 not being clear [24]. The number of experts or participants who assigned 3 points or higher was divided by the total number of people to obtain the content validity index (CVI); a CVI of ≥0.8 indicated that the item had good content validity [24,25].

To check whether people understood the instructions, response format, and items, as intended, cognitive interviews were undertaken [17]. Respondents were asked, "What were you thinking when you were answering that question?" This question was intended to elicit the cognitive process behind the answers. The following prompt was used, if needed: "Why did you select that response option?" Where relevant, items were adjusted, focusing on concepts related to health care and eHealth technologies.

Consensus Meeting

The meaning of items in the final translation was verified with the developers (authors LK and RHO) through written reflections on the back translation and a consensus meeting that also included a Mandarin-English bilingual and eHL expert (author CC) and the research team. The purpose was to confirm the forward translation and identify words, phrases, or concepts that were inconsistent with the item intents and implement revisions.

Phase 2. Psychometric Testing

The Chinese version of the eHLQ was then administered to people with chronic disease to explore its psychometric properties.

Recruitment

The participants were a random sample of Chinese adults attending the outpatient departments of the cardiology, nephrology, endocrinology, and family medicine units at several hospitals, such as medical centers, regional hospitals, and district hospitals, in towns or cities in Taiwan. To participate in this study, the inclusion criteria were (1) a diagnosis of type 2 diabetes mellitus, heart disease, or chronic kidney disease for more than 3 months; (2) the ability to clearly communicate in Mandarin and Taiwanese; and (3) age over 20 years. Diabetes, heart disease, and kidney disease were selected because they are the main focus of eHealth care in case management programs in Taiwan. For psychometric testing, a sample size of 300 was considered adequate [21,25].

Data Collection Procedures

Participants' data were collected by trained researchers. Case managers identified potentially eligible participants. At study inception, 1834 eligible people were identified on 2218 lists at 4 health services sites. Each person was randomly assigned a number, and 3 of every 10 (30%) were randomly selected by computer for inclusion in the study. This resulted in 550 (28.99%) people being selected, of which 442 (80.4%) participated. If a participant could not complete the questionnaire due to vision issues, such as presbyopia or myopia, an interviewer assisted the respondent. In the case of any hesitation from respondents during reading, research assistants simply repeated the item verbatim with no additional interpretation. Demographic information was collected, including age, gender, education, employment status, marital status, number of comorbidities, perceived health status in the previous month, income, and activities of daily living. Data were collected from October 2017 to February 2019.

Ethical Consideration

This study was approved by the institutional review board of the enrolling hospitals (2017-04-002CC; YM104135E2). Written informed consent was obtained prior to data collection.

Data Analysis

Data analysis was performed using SPSS Statistics version 23 (IBM Corporation, Armonk, NY, USA), STATA version 15.1 (Stata Corporation, College Station, TX, USA), and Mplus Version 8.3 (Muthén & Muthén, Los Angeles, CA, USA) [26]. For demographic characteristics, continuous variables were assessed by means and SD, whereas categorical variables were reported by frequency and percentage. Cronbach α was used to estimate internal consistency, with α≥.7 indicating acceptable reliability [25,27].

To check the assumptions of unidimensionality and local independence for IRT testing, a 1-factor model confirmatory factor analysis (CFA) for each of the 7 scales was fitted to the data using the weighted least squares mean and variance adjusted (WSLMV) estimation available in Mplus. The fit indices comparative fit Index (CFI) and standardized root mean residual (SRMR) were examined based on the 2-index strategy recommended by Shi et al [28] for models with small degrees of freedom. Although the chi-square test is a commonly used fit index, it has been found that it is sensitive to sample size and always rejects models when the sample size is large, while severe deviations from normality may also lead to model rejections [29,30]. The other commonly used fit index root-mean-square error of approximation (RMSEA) was also not an appropriate index for this study as the RMSEA has been found to reject models with small degrees of freedom and using it for assessment can be problematic [28,31]. Indication of a close fit for the CFI was >0.95 and for the SRMR was ≤0.08 [32]. Further investigation of local independence was by inspection of standardized factor loadings, modification indices, and standardized expected parameter change (SEPC) generated in the Mplus output. It has been recommended to use both of

these statistics to examine any model misspecification, with a large modification index combined with a positive value of SEPC >0.20 indicating misspecification [33,34].

For IRT analysis, the generalized partial credit model available in STATA version 15.1 was used to estimate item characteristic curves (ICCs), which describe the relationship between a respondent's ability and how they would respond to an item [35]. The two parameters of the ICC, item discrimination and item location/difficulty, were also evaluated. Item discrimination can detect subtle differences in the respondents' abilities, and a steeper slope of the ICC indicates a higher discrimination of the item [36,37], while item difficulty shows where the item functions best along the trait scale [35].

Known-group validity was evaluated by exploring associations between eHLQ scales and educational level with one-way ANOVA and Schaffer post hoc testing. $P<.05$ indicated statistical significance. The scales were tested for Gaussian distribution prior to ANOVA. All scales exhibited a normal distribution. Based on a review of previous eHL studies, it would be expected that the educational level would be positively related with all eHLQ scales (ie, lower education would be associated with lower eHLQ scores), although some studies indicate that feeling safe and in control may be inversely related or not related to eHL [38-41]. For the purpose of this study, education was aggregated to 6 International Standard Classification of Education 2011 (ISCED-2011) levels [42]:

- Lower than primary school equivalent to ISCED-2011 levels 0 and 1
- Junior high school equivalent to ISCED-2011 level 2
- Senior high school equivalent to ISCED-2011 level 3
- College equivalent to ISCED-2011 levels 4 and 5
- University equivalent to ISCED-2011 level 6
- Graduate school equivalent to ISCED-2011 levels 7 and 8

Results

Demographics and eHealth Literacy Scores

A total of 420 people who met the inclusion criteria completed the questionnaire in full. The response rate was 442 of 550 (80.4%) participants. Reasons for nonparticipation included lack of time, feeling of fatigue, and disinterest. Only 20 of 442 (4.5%) participants were excluded due to missing data.

The participants' mean (SD) age was 54.7 (13.1) years (range 25-89 years), 280 of 420 (66.7%) were between 50 and 64 years old, the majority (259/420, 61.7%) were male, 136 of 420 (32.4%) had completed junior high school or below, and 219 of 420 (52.1%) were unemployed. The monthly income in the previous year was below New Taiwan dollar (NTD) 20,000 (approximately US $700; the NTD-USD exchange rate was of 2020) for 247 of 420 (58.8%) participants. Regarding health status, most were living with 2 or more chronic diseases (Table 1).

The mean scores of the eHLQ scales ranged from 2.37 to 3.08. Respondents reported the highest scores on scale 2 (understanding of health concepts and language) and the lowest scores on scale 6 (access to digital services that work). See Table 2.

Table 1. Characteristics of participants (N=420).

Characteristics	Participants
Age (years), n (%)	
≤50	58 (13.8)
51-64	139 (33.1)
65-74	141 (33.6)
≥75	82 (19.5)
Gender, n (%)	
Male	259 (61.7)
Female	161 (38.3)
Education, n (%)	
Lower than primary school (ISCED[a]-2011 levels 0 and 1)	82 (19.5)
Junior high school (ISCED-2011 level 2)	54 (12.9)
Senior high school (ISCED-2011 level 3)	69 (16.4)
College (ISCED-2011 levels 4 and 5)	117 (27.9)
University (ISCED-2011 level 6)	65 (15.5)
Graduate school (ISCED-2011 levels 7 and 8)	33 (7.9)
Employment status, n (%)	
Not working (retired or unemployed)	219 (52.1)
Working	201(47.9)
Marital status, n (%)	
Single	147 (35.0)
Married or with a partner	273 (65.0)
Chronic disease, n (%)[b]	
Diabetes mellitus	96 (22.9)
Hypertension	249 (59.3)
Cardiovascular disease	261 (62.1)
Hyperlipidemia	104 (24.8)
Chronic kidney disease	175 (41.7)
Health status, n (%)	
1 disease	53 (12.7)
2 diseases	198 (47.1)
≥3 diseases	169 (40.2)
Monthly income in the last year, n (%)	
<NTD[c] 20,000 (~US $700) low	247 (58.8)
NTD 20,000-40,000 (~US $700-1400) lower middle	74 (17.6)
NTD 40,001-60,000 (~US $700-1800) middle	43 (10.2)
>NTD 60,000 (~>US $1800) middle higher	56 (13.3)
Living status, n (%)	
Lived alone	47 (11.2)
With spouse	227 (54.1)
With children	146 (34.8)

[a]ISCED: International Standard Classification of Education.

[b]More than 1 response was possible.

[c]NTD: New Taiwan dollar.

Table 2. The eHealth Literacy Questionnaire (eHLQ) scale scores and internal consistency.

Scale number	Scale name	Mean (SD)	Cronbach α
1	Using technology to process health information	2.41 (0.95)	.95
2	Understanding of health concepts and language	3.08 (0.57)	.75
3	Ability to actively engage with digital services	2.45 (0.91)	.90
4	Feel safe and in control	2.73 (0.78)	.87
5	Motivated to engage with digital services	2.49 (0.96)	.93
6	Access to digital services that work	2.37 (0.84)	.91
7	Digital services that suit individual needs	2.45 (0.99)	.90

Phase 1. Translation

Ten experts reviewed the item content validity, which resulted in scale CVIs from 0.88 to 0.95, well above the acceptable level of 0.80. For the item-level CVI, the lowest score was 0.80 (items 3, 17, 18, and 34), which was acceptable.

A total of 45 people participated in a cognitive interview. Most interviewees suggested some words or terms that required more description for clarity and ease of understanding. For instance, they were unsure of terms covering health technology services, people who required health information, and authorized people. In addition, they did not easily link these terms to their disease management situation and the relevant health care system. The instructions and definitions of terms in the questionnaire introduction were therefore revised. All other items were understood as intended and no further changes made. The interview process took 10-15 minutes to complete.

The consensus meeting identified 4 items that required minor refinement (items 6, 19, 24, and 25). For example, for items 19 and 24, the word “find” was originally translated as “發現.” However, the consensus meeting revealed that “find” includes the meaning “perceive” and “believe” and so was replaced with “發覺” to reflect the intended meaning. See Multimedia Appendix 2 for the final Chinese version of the eHLQ.

Phase 2. Psychometric Testing

Reliability

The internal consistency coefficients are shown in Table 2. All scales had α>.80 except scale 2, which had α=.75.

Known-Group Validity

There were striking differences among the educational levels, with a clear monotonic increase in scores for all scales from the lowest to the highest education, except for the higher levels of education for scale 2. A comparison of the 7 scales of the Chinses version of the eHLQ across educational levels is shown in Multimedia Appendix 3. The largest differences between the lowest and highest education levels were for scale 6, where people with the lowest education, on average, scored 1.08, indicating that almost all respondents strongly disagreed that they could access technologies that worked. The smallest differences in education were seen for scale 2, where the average score of the lowest-education group was 2.89 compared with 3.24 in the highest-education group.

Construct Validity

The 1-factor CFA models generally fitted the data well on all scales based on the CFI and SRMR fit indices, and the SEPC values were below 0.2 except for scale 2, with CFI=0.95, SRMR=0.04, the largest modification index=86.2 for eHLQ26 (“I use measurements about my body to help me understand my health”) and eHLQ15 (“I understand medical results about me”), and SEPC=0.29. This finding indicates that content within these 2 particular items is related in a unique way, in addition to how they are related to the latent variable of the scale’s construct. A model with a correlated residual between these 2 items was tested, and the results demonstrated a close fit, with no large modification index or SEPC (see Multimedia Appendix 4). Standardized factor loadings were significant for all scales, with loadings >.50 (Multimedia Appendix 5). As such, the unidimensionality and local independence assumptions were met, except for scale 2, for which the local independence assumption might not hold and the IRT results for this scale needed to be interpreted with caution.

Item Response Theory

IRT analysis showed that respondents could use the response options in a consistent way and that no items were found to have disordered thresholds. See Multimedia Appendix 6 for the ICCs of the Chinese version of the eHLQ. Inspection of the steepness of the slopes of the ICCs and the estimated item discrimination parameters showed that all items except items 15 and 26 of scale 2 had acceptable-to-good discrimination between people with different levels of ability. The estimated item difficulty parameters demonstrated a range of difficulty levels within each scale except scale 7 (digital services that suit individual needs). The widest difficulty range was noted for scale 4 (feel safe and in control; range 0.82-1.59) and scale 6 (range 0.82-1.59). Scale 7 had the narrowest range (0.70-0.79). However, all results within the scales were statistically significantly different (Table 3).

Table 3. Item response theory (IRT) analysis of the Chinese version of the eHealth Literacy Questionnaire (eHLQ) using the generalized partial credit model.

Scale item	Item difficulty (95% CI)	Item discrimination (95% CI)
1. Using technology to process health information		
eHLQ7	0.69 (0.57-0.81)	7.48 (6.08-8.89)
eHLQ11	0.84 (0.70-0.97)	5.73 (4.67-6.79)
eHLQ13	0.76 (0.64-0.89)	6.53 (5.34-7.72)
eHLQ20	0.89 (0.73-1.02)	5.21 (4.24-6.19)
eHLQ25	1.02 (0.85-1.18)	4.30 (3.50-5.09)
2. Understanding of health concepts and language		
eHLQ5	0.92 (0.67-1.17)	1.18 (0.93-1.44)
eHLQ12	1.19 (0.89-1.48)	1.08 (0.85-1.30)
eHLQ15	0.83 (0.22-1.44)	0.38 (0.22-0.53)
eHLQ21	1.40 (1.01-1.80)	0.78 (0.60-0.96)
eHLQ26	1.05 (0.13-1.97)	0.27 (0.11-0.44)
3. Ability to actively engage with digital services		
eHLQ4	0.82 (0.69-0.96)	5.10 (4.16-6.04)
eHLQ6	0.76 (0.66-0.92)	6.06 (4.96-7.17)
eHLQ8	0.79 (0.66-0.94)	4.28 (3.49-5.08)
eHLQ17	1.09 (0.76-1.34)	5.32 (4.34-6.30)
eHLQ32	0.84 (0.70-0.97)	5.89 (4.80-6.97)
4. Feel safe and in control		
eHLQ1	0.82 (0.56-1.09)	1.05 (0.83-1.27)
eHLQ10	1.59 (1.05-1.80)	0.98 (0.77-1.19)
eHLQ14	1.25 (0.95-1.54)	1.24 (0.99-1.49)
eHLQ22	0.93 (0.67-1.19)	1.12 (0.89-1.35)
eHLQ30	1.13 (0.84-1.42)	1.09 (0.87-1.32)
5. Motivated to engage with digital services		
eHLQ2	0.68 (0.64-0.91)	4.41 (3.59-5.23)
eHLQ19	0.59 (0.51-0.80)	9.55 (7.73-11.36)
eHLQ24	0.85 (0.70-0.96)	7.30 (5.93-8.68)
eHLQ27	0.80 (0.68-0.95)	5.83 (4.75-6.92)
eHLQ35	0.77 (0.64-0.90)	5.11 (4.18-6.05)
6. Access to digital services that work		
eHLQ3	1.67 (1.33-2.02)	1.26 (1.01-1.51)
eHLQ9	0.83 (0.70-0.97)	5.38 (4.39-6.37)
eHLQ16	1.10 (0.90-1.33)	2.22 (1.81-2.63)
eHLQ23	0 .90 (0.75-1.05)	4.17 (3.40-4.94)
eHLQ29	1.12 (0.91-1.32)	2.56 (2.08 -3.03)
eHLQ34	0.89 (0.75-1.05)	4.44 (3.60-5.28)
7. Digital services that suit individual needs		
eHLQ18	0.71 (0.65-0.89)	12.15 (9.58-14.73)
eHLQ28	0.77 (0.65-0.89)	9.60 (7.72-11.47)
eHLQ31	0.79 (0.67-0.91)	10.51 (8.39-12.63)

Scale item	Item difficulty (95% CI)	Item discrimination (95% CI)
eHLQ33	0.73 (0.61-0.85)	10.63 (8.52-12.73)

Discussion

Principal Findings

This study undertook a rigorous process of translating the eHLQ, ensuring cultural appropriateness for the Chinese context, and examined several key indicators of validity based on data derived from a large randomly selected sample of people with chronic conditions from diverse demographic backgrounds. In this setting, the eHLQ was found to have strong-to-acceptable psychometric properties using both classical test and IRT approaches. The translated and culturally adapted eHLQ items were found to be highly coherent with the original intended meanings and psychometric properties.

For a translated version to be considered a robust questionnaire in this setting, not only are systematic and standardized translation processes required but also a verification process [16,20,21,43,44]. Our translation process provides evidence for a validity argument of the translated version, as recommended by Hawkins et al [19,20,43]. The Translation Integrity Procedure with the detailed construct and item intent descriptions supported this process with a common foundation for the translation team to negotiate the nuances of item meanings to maximize construct equivalence, minimize threats to construct validity during the translation process, and generate qualitative validity evidence for score interpretation and use in a new linguistic context [20]. Overall, the process ensured that the translated version had semantic, idiomatic, experiential, and conceptual equivalence with the original [21,45].

All of the items within the scales loaded strongly on their respective factors. With 1 modification, all 1-factor models fitted the data well. For scale 2, a correlated residual was added between 2 items, which may have been independently related due to the hospital setting, where the content of both items related to medical results and measurement about one's body (items 15 and 26, respectively). Cronbach α, which is frequently inflated due to excess items within a scale, was >.85 for all scales except Scale 2, which still had an acceptable reliability of .75. Consistent with the validity-driven approach, the development of the eHLQ ensured that a minimal number of items (4-6 for the eHLQ scales) were carefully generated to capture the full breadth of each individual construct. This ambitious parsimonious constraint was reproduced in the Mandarin language setting, which included people with diverse educational levels and health conditions. The original Danish validity-testing study reported similar internal consistency (range .77-.86) [17], where scale 2 also had the lowest value. Of note, the internal consistency of 3 scales (1, 5, and 6) was greater than .90, which may indicate some translated items may be understood in an overly similar way in this Chinese population [46]. In addition, almost 60% of participants used case management services and had the same experience in eHealth care in using mobile health monitors, such as blood sugar and blood pressure, so it is possible that most respondents had similar experiences relative to multiple items within scales. Importantly, for all the scales except scale 7, a range of difficulties was found, indicating that differences among individuals were expressed across items.

Our study showed that the average scores of people with primary or lower education is substantially lower than those of other more educated groups. The findings were striking, especially for scale 6, where the average score was 1.08, which indicates that almost all respondents with primary or lower education responded "strongly disagree" on all items in this scale. In contrast, for the highest-education group, most respondents indicated they "agreed", with a score of 2.93, indicating they had access to digital services that work. Similar patterns were also seen for scales 1, 3, 5, and 7; however, the lowest-education group, on average, on all items per scale marked "disagree" rather than "strongly disagree." These findings demonstrate a stark social gradient related to education, where people with higher education clearly have higher eHL than those with lower education. This is generally in line with other findings that lower education is associated with lower eHL [9,38-40]. In a study of nursing students in Denmark, which also used the eHLQ, found that graduate-level students scored higher than entry-level nursing students on scales 1-3, with no differences on the other scales [47]. The eHLQ appears to be a promising tool to understand digital access to different educational categories and is therefore likely to be a useful tool for understanding socioeconomic determinants of digital access inequity.

IRT analysis also provided insights into the psychometric properties of the eHLQ. This analysis showed good sensitivity in detecting participants with different levels of ability, as well as representing a range of difficulty within each scale, echoing the findings of the initial development studies in Denmark [17]. Only 2 items relating to medical results and measurements, items 15 and 26 of scale 2, were found to have low item discrimination. These items were also found to be problematic in classical test theory analysis (reliability and CFA, as noted in the Construct Validity section). In this health care setting, respondents may be prone to providing socially desirable answers, common in Chinese culture. People tend to be cautious to ask and share problems with health care professionals to avoid discrimination [48,49], or they may tend to hide a lack of understanding about their tests or measurements. Although it is possible that the performance of these items relates to the particular characteristics of the respondents in our study, further work in different populations alongside linguistic evaluation will shed light on these items.

Strengths

An important strength of this study was the heterogeneity of the sample. People with a range of chronic diseases from various hospitals and clinics were included. These are key settings and populations for the application of the eHLQ in future studies, and given that overall good psychometric properties were observed in this diverse sample, it is likely to be a robust measure in other related settings.

Limitations

The CFA results indicated that the local independence assumption for scale 2 might not have been met; therefore, the IRT results for this scale need to be interpreted with caution. This study relied on self-report, and therefore, the respondents' answers may be prone to recall bias and social desirability, similar to other self-report measures. Some participants completed the questionnaire with the help of a research assistant. This only occurred on 42 occasions, and although the administration mode may affect the respondents' scores, this was regarded as an important process to ensure people with low eHL were included to maximize the representativeness of the sample. Future research should explore whether the administration mode introduces bias.

Conclusion

This study demonstrates that the eHLQ has good linguistic equivalence and psychometric properties, following a rigorous translation and cultural adaptation process and extensive psychometric testing using both classical test theory and IRT approaches. The 7 scales of the eHLQ can efficiently assess diverse dimensions of eHL of people across chronic diseases. The questionnaire is likely to enable health care providers and eHealth system developers to better understand people's ability to engage with and use technology so that these systems can be developed, evaluated, and redesigned to meet the health and equity needs of their communities. As such, the eHLQ can be used as a reference to design adaptive care programs to improve the quality and effectiveness of care. This may also help avoid the health disparities created by the advancement of digital technologies.

Acknowledgments

The authors thank all the patients, case managers, and research assistants (Yu-Shan Huang, I-Ying Lu, Man-I Chao) for their assistance in the study process. We would also like to thank Professor Gerald Elsworth for his valuable advice on psychometric testing. This work was funded by the Ministry of Science and Technology, Taiwan (MOST 105-2314-B-010-039), and the Yen Tjing Ling Medical Foundation, Taiwan (CI-106-29). RHO was funded, in part, through a National Health and Medical Research Council of Australia Principal Research Fellowship (#APP1155125).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Scale names and construct definitions of the eHealth Literacy Questionnaire (eHLQ).
[DOCX File , 29 KB - jmir_v24i1e32855_app1.docx]

Multimedia Appendix 2
Scales of the Chinese version of the eHealth Literacy Questionnaire (eHLQ).
[DOCX File , 29 KB - jmir_v24i1e32855_app2.docx]

Multimedia Appendix 3
Comparison of the 7 scales of the Chinese version of the eHealth Literacy Questionnaire (eHLQ) across educational levels.
[DOCX File , 45 KB - jmir_v24i1e32855_app3.docx]

Multimedia Appendix 4
Model fit indices for the 1-factor confirmatory factor analysis of the Chinese version of the eHealth Literacy Questionnaire (eHLQ).
[DOCX File , 30 KB - jmir_v24i1e32855_app4.docx]

Multimedia Appendix 5
Standardized factor loadings of the 7 one-factor models of the Chinese version of the eHealth Literacy Questionnaire (eHLQ).
[DOCX File , 36 KB - jmir_v24i1e32855_app5.docx]

Multimedia Appendix 6
Item characteristic curves (ICCs) of the Chinese version of the eHealth Literacy Questionnaire (eHLQ).
[DOCX File , 840 KB - jmir_v24i1e32855_app6.docx]

References

1. Norman CD, Skinner HA. eHealth literacy: essential skills for consumer health in a networked world. J Med Internet Res 2006 Jun 16;8(2):e9 [FREE Full text] [doi: 10.2196/jmir.8.2.e9] [Medline: 16867972]

2. Gilstad H. Toward a comprehensive model of eHealth literacy. 2013 Mar 11 Presented at: Proceedings of the 2nd European Workshop on Practical Aspects of Health Informatics; 2013/03/11; Trondheim, Norway p. 1-11 URL: http://ceur-ws.org2014/
3. Monkman H, Griffith J, Kushniruk A. Evidence-based heuristics for evaluating demands on eHealth literacy and usability in a mobile consumer health application. Stud Health Technol Inform 2015;216:358-362. [Medline: 26262071]
4. Chung S, Nahm E. Testing reliability and validity of the eHealth Literacy Scale (eHEALS) for older adults recruited online. Comput Inform Nurs 2015 Apr;33(4):150-156 [FREE Full text] [doi: 10.1097/CIN.0000000000000146] [Medline: 25783223]
5. Norgaard O, Furstrand D, Klokker L, Karnoe A, Batterham R, Kayser L. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manag E-Learn Int J 2015 Dec 26;7(4):522-540 [FREE Full text] [doi: 10.34105/j.kmel.2015.07.035]
6. Kreps GL, Neuhauser L. New directions in eHealth communication: opportunities and challenges. Patient Educ Couns 2010 Mar;78(3):329-336. [doi: 10.1016/j.pec.2010.01.013] [Medline: 20202779]
7. Knight E, Stuckey MI, Petrella RJ. Health promotion through primary care: enhancing self-management with activity prescription and mHealth. Phys Sportsmed 2014 Sep;42(3):90-99 [FREE Full text] [doi: 10.3810/psm.2014.09.2080] [Medline: 25295771]
8. Alpay L, van der Boog P, Dumaij A. An empowerment-based approach to developing innovative e-health tools for self-management. Health Informatics J 2011 Dec;17(4):247-255 [FREE Full text] [doi: 10.1177/1460458211420089] [Medline: 22193825]
9. Foley K, Freeman T, Ward P, Lawler A, Osborne RH, Fisher M. Exploring access to, use of and benefits from population-oriented digital health services in Australia. Health Promot Int 2021 Aug 30;36(4):1105-1115. [doi: 10.1093/heapro/daaa145] [Medline: 33367568]
10. Aponte J, Nokes KM. Electronic health literacy of older Hispanics with diabetes. Health Promot Int 2017 Jun 01;32(3):482-489 [FREE Full text] [doi: 10.1093/heapro/dav112] [Medline: 26681770]
11. Diamantidis CJ, Becker S. Health information technology (IT) to improve the care of patients with chronic kidney disease (CKD). BMC Nephrol 2014 Jan 09;15(7):7-6 [FREE Full text] [doi: 10.1186/1471-2369-15-7] [Medline: 24405907]
12. Kayser L, Kushniruk A, Osborne RH, Norgaard O, Turner P. Enhancing the effectiveness of consumer-focused health information technology systems through eHealth literacy: a framework for understanding users' needs. JMIR Hum Factors 2015 May 20;2(1):e9 [FREE Full text] [doi: 10.2196/humanfactors.3696] [Medline: 27025228]
13. Neter E, Brainin E. eHealth literacy: extending the digital divide to the realm of health information. J Med Internet Res 2012 Jan 27;14(1):e19 [FREE Full text] [doi: 10.2196/jmir.1619] [Medline: 22357448]
14. Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. J Med Internet Res 2006 Nov 14;8(4):e27 [FREE Full text] [doi: 10.2196/jmir.8.4.e27] [Medline: 17213046]
15. Karnoe A, Kayser L. How is eHealth literacy measured and what do the measurements tell us? A systematic review. Knowl Manag E-Learn Int J 2015 Dec 26;7(4):576-600 [FREE Full text] [doi: 10.34105/j.kmel.2015.07.038]
16. Buchbinder R, Batterham R, Elsworth G, Dionne CE, Irvin E, Osborne RH. A validity-driven approach to the understanding of the personal and societal burden of low back pain: development of a conceptual and measurement model. Arthritis Res Ther 2011;13(5):R152 [FREE Full text] [doi: 10.1186/ar3468] [Medline: 21933393]
17. Kayser L, Karnoe A, Furstrand D, Batterham R, Christensen KB, Elsworth G, et al. A multidimensional tool based on the ehealth literacy framework: development and initial validity testing of the eHealth Literacy Questionnaire (eHLQ). J Med Internet Res 2018 Feb 12;20(2):e36 [FREE Full text] [doi: 10.2196/jmir.8371] [Medline: 29434011]
18. Cheng C, Elsworth G, Osborne RH. Validity evidence based on relations to other variables of the eHealth Literacy Questionnaire (eHLQ): Bayesian approach to test for known-groups validity. J Med Internet Res 2021 Oct 14;23(10):e30243 [FREE Full text] [doi: 10.2196/30243] [Medline: 34647897]
19. Hall DA, Zaragoza Domingo S, Hamdache LZ, Manchaiah V, Thammaiah S, Evans C, International Collegium of Rehabilitative AudiologyTINnitus Research NETwork. A good practice guide for translating and adapting hearing-related questionnaires for different languages and cultures. Int J Audiol 2018 Mar;57(3):161-175 [FREE Full text] [doi: 10.1080/14992027.2017.1393565] [Medline: 29161914]
20. Hawkins M, Cheng C, Elsworth GR, Osborne RH. Translation method is validity evidence for construct equivalence: analysis of secondary data routinely collected during translations of the Health Literacy Questionnaire (HLQ). BMC Med Res Methodol 2020 May 26;20(1):130 [FREE Full text] [doi: 10.1186/s12874-020-00962-8] [Medline: 32456680]
21. Sousa VD, Rojjanasrirat W. Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: a clear and user-friendly guideline. J Eval Clin Pract 2011 Apr;17(2):268-274 [FREE Full text] [doi: 10.1111/j.1365-2753.2010.01434.x] [Medline: 20874835]
22. Epstein J, Osborne RH, Elsworth GR, Beaton DE, Guillemin F. Cross-cultural adaptation of the Health Education Impact Questionnaire: experimental study showed expert committee, not back-translation, added value. J Clin Epidemiol 2015 Apr;68(4):360-369 [FREE Full text] [doi: 10.1016/j.jclinepi.2013.07.013] [Medline: 24084448]
23. Hawkins M, Osborne R. Health Literacy Questionnaire (HLQ) Translation and Cultural Adaptation Procedure: Version 4. Australia: Swinburne University; 2019.

24. Rubio DM, Berg-Weger M, Tebb SS, Lee ES, Rauch S. Objectifying content validity: conducting a content validity study in social work research. Soc Work Res 2003 Jun 01;27(2):94-104 [FREE Full text] [doi: 10.1093/swr/27.2.94]
25. Polit D, Beck C. Essentials of Nursing Research: Appraising Evidence for Nursing Practice. 8th ed. Philadelphia, PA: Wolters Kluwer/Lippincott William & Wilkins; 2013:300-428.
26. Muthén LK, Muthén BO. Mplus (Mac) Version 8.3. Los Angeles, CA: Muthén & Muthén; 2017.
27. Priest HM. Essentials of nursing research: methods, appraisal, and utilization. Nurse Res 2006 Jul 01;13(4):91-92 [FREE Full text] [doi: 10.7748/nr.13.4.91.s11] [Medline: 27702222]
28. Shi D, DiStefano C, Maydeu-Olivares A, Lee T. Evaluating SEM model fit with small degrees of freedom. Multivariate Behav Res 2021 Feb 12;12:1-36. [doi: 10.1080/00273171.2020.1868965] [Medline: 33576257]
29. Kenny DA, McCoach DB. Effect of the number of variables on measures of fit in structural equation modeling. Struct Equ Mod 2009 Nov 19;10(3):333-351 [FREE Full text] [doi: 10.1207/S15328007SEM1003_1]
30. Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. Elecron J Bus Res Methods 2008 Jan 01:53-60 [FREE Full text]
31. Kenny DA, Kaniskan B, McCoach DB. The performance of RMSEA in models with small degrees of freedom. Sociol Methods Res 2014 Jul 24;44(3):486-507 [FREE Full text] [doi: 10.1177/0049124114543236]
32. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Struct Equ Model 1999 Jan;6(1):1-55 [FREE Full text] [doi: 10.1080/10705519909540118]
33. Saris WE, Satorra A, van der Veld WM. Testing structural equation models or detection of misspecifications? Struct Equ Model 2009 Oct 06;16(4):561-582 [FREE Full text] [doi: 10.1080/10705510903203433]
34. Whittaker TA. Using the modification index and standardized expected parameter change for model modification. J Exp Educ 2012 Jan;80(1):26-44 [FREE Full text] [doi: 10.1080/00220973.2010.531299]
35. Nguyen TH, Han H, Kim MT, Chan KS. An introduction to item response theory for patient-reported outcome measurement. Patient 2014;7(1):23-35 [FREE Full text] [doi: 10.1007/s40271-013-0041-0] [Medline: 24403095]
36. Zanon C, Hutz CS, Yoo H, Hambleton RK. An application of item response theory to psychological test development. Psicol Refl Crít 2016 Apr 18;29(1):1-10 [FREE Full text] [doi: 10.1186/s41155-016-0040-x]
37. Hambleton RK, Swaminathan H. Item Response Theory: Principles and Applications. The Netherlands: Springer Science & Business Media; 2013:62-68.
38. Liu P, Yeh L, Wang J, Lee S. Relationship between levels of digital health literacy based on the Taiwan Digital Health Literacy Assessment and accurate assessment of online health information: cross-sectional questionnaire study. J Med Internet Res 2020 Dec 21;22(12):e19767 [FREE Full text] [doi: 10.2196/19767] [Medline: 33106226]
39. Xesfingi S, Vozikis A. eHealth literacy: in the quest of the contributing factors. Interact J Med Res 2016 May 25;5(2):e16 [FREE Full text] [doi: 10.2196/ijmr.4749] [Medline: 27226146]
40. Richtering SS, Morris R, Soh S, Barker A, Bampi F, Neubeck L, et al. Examination of an eHealth literacy scale and a health literacy scale in a population with moderate to high cardiovascular risk: Rasch analyses. PLoS One 2017 Apr 27;12(4):e0175372 [FREE Full text] [doi: 10.1371/journal.pone.0175372] [Medline: 28448497]
41. Milne RA, Puts MTE, Papadakos J, Le LW, Milne VC, Hope AJ, et al. Predictors of high eHealth literacy in primary lung cancer survivors. J Cancer Educ 2015 Dec;30(4):685-692 [FREE Full text] [doi: 10.1007/s13187-014-0744-5] [Medline: 25355524]
42. International Standard Classification of Education (ISCED). Eurostat Statistic Explained. URL: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=International_Standard_Classification_of_Education_(ISCED) [accessed 2020-09-10]
43. Hawkins M, Elsworth GR, Osborne RH. Application of validity theory and methodology to patient-reported outcome measures (PROMs): building an argument for validity. Qual Life Res 2018 Jul;27(7):1695-1710 [FREE Full text] [doi: 10.1007/s11136-018-1815-6] [Medline: 29464456]
44. Hawkins M, Elsworth GR, Nolte S, Osborne RH. Validity arguments for patient-reported outcomes: justifying the intended interpretation and use of data. J Patient Rep Outcomes 2021 Jul 30;5(1):64 [FREE Full text] [doi: 10.1186/s41687-021-00332-y] [Medline: 34328558]
45. Sperber AD. Translation and validation of study instruments for cross-cultural research. Gastroenterology 2004 Jan;126(1 Suppl 1):S124-S128 [FREE Full text] [doi: 10.1053/j.gastro.2003.10.016] [Medline: 14978648]
46. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ 2011 Jun 27;2:53-55 [FREE Full text] [doi: 10.5116/ijme.4dfb.8dfd] [Medline: 28029643]
47. Holt KA, Overgaard D, Engel LV, Kayser L. Health literacy, digital literacy and eHealth literacy in Danish nursing students at entry and graduate level: a cross sectional study. BMC Nurs 2020 Apr 10;19:22 [FREE Full text] [doi: 10.1186/s12912-020-00418-w] [Medline: 32308559]
48. Mackert M, Mabry-Flynn A, Donovan EE, Champlin S, Pounders K. Health literacy and perceptions of stigma. J Health Commun 2019;24(11):856-864 [FREE Full text] [doi: 10.1080/10810730.2019.1678705] [Medline: 31630662]
49. Schillinger D. The intersections between social determinants of health, health literacy, and health disparities. Stud Health Technol Inform 2020 Jun 25;269:22-41 [FREE Full text] [doi: 10.3233/SHTI200020] [Medline: 32593981]

Abbreviations

CFA: confirmatory factor analysis
CFI: comparative fit index
CVI: content validity index
eHealth: electronic health
eHL: eHealth literacy
eHLF: eHealth Literacy Framework
eHLQ: eHealth Literacy Questionnaire
ICC: item characteristic curve
IRT: item response theory
ISCED: International Standard Classification of Education
RMSEA: root-mean-square error of approximation
SEPC: standardized expected parameter change
SRMR: standardized root mean residual

Edited by T Leung; submitted 12.08.21; peer-reviewed by A Tannoubi, S Wei, H Salim, R Haase; comments to author 27.09.21; accepted 31.12.21; published 19.01.22.

Please cite as:
Chen YC, Cheng C, Osborne RH, Kayser L, Liu CY, Chang LC
Validity Testing and Cultural Adaptation of the eHealth Literacy Questionnaire (eHLQ) Among People With Chronic Diseases in Taiwan: Mixed Methods Study
J Med Internet Res 2022;24(1):e32855
URL: https://www.jmir.org/2022/1/e32855
doi:10.2196/32855
PMID:35044310

Original Paper

The Gap Between Self-Rated Health Information Literacy and Internet Health Information-Seeking Ability for Patients With Chronic Diseases in Rural Communities: Cross-sectional Study

Zhuoxin Wang[1*], MSc; Yanyan Fan[1*], MSc; Hekai Lv[2,3*], MSc; Shanshan Deng[1*], MSc; Hui Xie[1*], PhD; Li Zhang[1*], MSc; Aijing Luo[4,5*], PhD; Fuzhi Wang[2,3*], PhD

[1]School of Nursing, Bengbu Medical College, Bengbu, China

[2]School of Health Management, Bengbu Medical College, Bengbu, China

[3]Innovation Team of Health Information Management and Application Research (BYKC201913), Bengbu Medical College, Bengbu, China

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

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

[*]all authors contributed equally

Corresponding Author:
Fuzhi Wang, PhD
School of Health Management
Bengbu Medical College
2600# Donghai Rd, Bengbu, Anhui Province, PRC
Bengbu, 233030
China
Phone: 86 18855202156
Email: wfz.bbmc@foxmail.com

Abstract

Background: The internet has become one of the most important channels for residents to seek health information, particularly in remote rural areas in China.

Objective: In this study, we aimed to explore the gap between self-rated health information literacy and internet health information seeking ability for patients with chronic diseases in rural communities and to preliminarily evaluate their barriers when seeking health information via the internet.

Methods: Residents from rural communities near Bengbu City and with chronic diseases were included in this study. A self-rated questionnaire was used to evaluate their health information literacy, 3 behavioral competency tasks were designed to preliminarily evaluate their ability to seek health information on the internet and semistructured interviews were used to investigate their barriers to obtaining health information via the internet. A small audiorecorder was used to record the interview content, and screen-recording software was used to record the participants' behavior during the web-based operational tasks.

Results: A total of 70 respondents completed the self-rated health information literacy questionnaire and the behavioral competence test, and 56 respondents participated in the semistructured interviews. Self-rated health information literacy (score out of 70: mean 46.21, SD 4.90) of the 70 respondents were moderate. Although 91% (64/70) of the respondents could find health websites, and 93% (65/70) of the respondents could find information on treatment that they thought was the best, 35% (23/65) of respondents did not know how to save the results they had found. The operational tasks indicated that most articles selected by the respondents came from websites with encyclopedic knowledge or answers from people based on their own experiences rather than authoritative health information websites. After combining the results of the semistructured interviews with the DISCERN scale test results, we found that most interviewees had difficulty obtaining high-quality health information via the internet.

Conclusions: Although the health information literacy level of patients with rural chronic disease was moderate, they lack the ability to access high-quality health information via the internet. The vast majority of respondents recognized the importance of accessing health information but were not very proactive in accessing such information.

(J Med Internet Res 2022;24(1):e26308) doi:10.2196/26308

KEYWORDS

online; health information; barriers to acquisition; middle-aged patients with chronic diseases; rural community; chronic conditions; chronic; rural; literacy; information seeking

Introduction

Chronic diseases have become major challenges to global health [1]. According to 2018 data from the World Health Organization, chronic diseases cause 41 million deaths every year, accounting for 71% of all deaths worldwide, and of these, 15 million occur among people aged 30 to 69 years [2]. China has documented significant decreases in the age of patients with chronic disease, and the number of such patients is increasing over time [3]. Previous studies [4] have found that health professionals are the main source of health knowledge for patients with chronic diseases. With the rapid development of information technology in recent decades, an increasing number of patients with chronic diseases choose to obtain health information through the internet [5,6], particularly in the vast rural areas of China [7]. Relevant studies [8] have shown that health-related interventions implemented via the internet can improve the health status of patients with chronic physical diseases. However, numerous studies have shown that the quality of internet health information is not optimal in China [9] or in other countries around the world [10,11]. Additionally, the low level of information literacy among the public in rural areas [12] causes some users to have difficulty in effectively using health information from the internet to improve their health statuses.

In the past decade, some studies [13,14] have suggested that health information literacy should be a key part of public health promotion in China. Health information literacy is the set of abilities that users have to recognize the need for health information, identify possible sources of information, and use those sources to retrieve relevant information; evaluate the quality of information and its applicability to a given situation; and analyze, understand, and use this information to make sound health decisions [15]. Despite the increasing amount of health information available on the internet, patients with limited health knowledge may lack the necessary skills to access the internet or make use of this information. A study [6] examining the relationship between eHealth literacy and health information–seeking behaviors and participation in mobile health research among African Americans found that most participants scored high on eHealth literacy but lacked the ability to distinguish between high- and low-quality health resources on the internet and to use internet information to make health decisions. The study also reported that people with lower education levels were less likely to use the internet to obtain health-related information [6]. Another cross-sectional study [16] found that patients with chronic diseases rely on health care professionals for health information regardless of their level of health literacy and that patients with low levels of health information literacy lacked the ability to use the internet to look up health information (ie, they had low health information–seeking behaviors on the internet). Therefore, in the context of the rapid development of global information technology, improving the public's ability to access quality health information through the internet remains an important element of health information literacy promotion.

The term *internet health information–seeking behavior* refers to the process in which users search for health knowledge or information on the internet to meet their own health information needs and to reduce the uncertainty of their health statuses [17]. A review [18] of existing studies on patients' internet health information–seeking and its impact on doctor-patient relationships and found that such information seeking could improve doctor–patient relationships. Another study [19] evaluated the characteristics of different types of internet users in seeking web-based health information and found that there were differences in the use of and access to health information among people of different ages, races, and socioeconomic status. Another study [20] found that a large proportion of older adult patients with chronic diseases use the internet to seek health information; the onset time and type of chronic diseases may play an important role in their internet health information–seeking behavior. At the same time, Latino individuals are less likely than white individuals to search for health information but are more likely to use health information to treat disease whereas African American individuals are more likely to use health information to maintain their health [21]. A couple studies in China [22,23] have been conducted on the internet health information–seeking behavior of patients with chronic diseases. A cross-sectional study [22] of 313 hospitalized patients with chronic diseases in Shanghai found that the patients' attitudes toward health information–seeking were in the middle to upper levels and the patients had a high demand for health information. Another study [23] investigated the health information–seeking behavior of hypertensive patients in Guizhou Province, and the results indicated that the health information needs of hypertensive patients were diverse. The main health information access channels were traditional interpersonal relationship channels (medical personnel), while new media network technology was seldom used; the digital divide is the main cause for this problem. We found that the current research methods for patients with chronic disease in China internet health information–seeking behavior are limited to questionnaire surveys and subjective evaluations. There is a lack of research employing interviews and operational evaluations of the internet health information–seeking behavior of patients with chronic diseases. The lack of such research suggested that we should conduct in-depth evaluations on the barriers faced by patients with chronic diseases in seeking health information on the internet from an objective and extensive research perspective.

In this study, questionnaires, semistructured interviews, and behavioral competency tests were used to explore the gap between self-rated health information literacy and internet health information–seeking abilities for patients with chronic diseases and to preliminarily evaluate their barriers to seeking health information via the internet. These results are expected to provide support for the government and health education

institutions to perform internet health information behavioral intervention for patients with chronic diseases in the future.

Methods

Research Tasks

Overview

This research investigation included 3 tasks. First, a self-rated health information literacy questionnaire was used to evaluate the health information literacy levels of the respondents. Second, 3 operational tasks were designed and administered to test the respondents' abilities to seek internet health information on a computer or mobile device. Finally, interviews on internet health information–seeking (containing 9 questions) were performed. Before the survey began, 3 experts (doctoral degrees and ample research experience) were invited to evaluate the validity of the survey materials used in this study, and appropriate modifications were made based on the experts' opinions (Multimedia Appendix 1).

Task 1: Self-Rated Health Information Literacy Questionnaire

A self-rated questionnaire was used to assess the health information literacy of respondents. The questionnaire was synthesized and revised from the 10-item Everyday Health Information Literacy scale [24] and modified in 2018 [25]. The scale contains 14 items divided into 4 dimensions: health information consciousness, health information–seeking, health information evaluation, and health information application. A higher score indicates a higher health information literacy level. The questionnaire has been widely used for assessing the health information literacy of Chinese digital immigrants in rural communities [12] and Chinese residents [25] and has shown good reliability, validity, and adaptability for the Chinese population.

Task 2: Internet Health Information-Seeking Behavior Ability Tests

Without receiving guidance from the researchers administering the test, the respondents were asked to log on to the internet to search for health information on their own, and respondents' behavior throughout the entire test was recorded with screen-recording software. In this task, the respondents were asked to complete 3 operational tasks: (1) use the internet to find what they thought was the best health website containing knowledge about chronic diseases; (2) according to their own health conditions, choose and save 2 to 3 articles related to health knowledge they considered valuable; and (3) choose an article that they thought discussed the best treatment for a certain chronic disease and save it (if the respondents had difficulty saving the article, the investigator helped them save it for subsequent analyses).

The DISCERN scale was used to evaluate the quality of the chronic disease treatment articles selected internet by the respondents. The DISCERN scale, developed by the British Library and the University of Oxford [26], is divided into 2 parts. The first part consists of 8 items and has a total score of 40; this part was used to evaluate the reliability of the website. The second part includes 7 items and has a total score of 35; this part was used to test the quality of the articles.

Task 3: Semistructured Interviews

The semistructured interviews focused on 2 topics: the retrieval process and strategy and the evaluation of the quality of health information on the internet. The investigator conducted an in-depth exchange with the respondents to understand their criteria for selecting the health information websites, their methods for identifying the quality of information and the difficulties they encountered in obtaining internet health information. The interviews were recorded throughout.

Participants

The participants of this study were patients with chronic diseases in Xiaobengbu Town, Bengbu City, Anhui Province. Two-stage sampling was used. In the first stage, 10 rural communities in the town of Xiaobengbu were identified using simple random sampling; in the second stage, a general practitioner from the community health service station was invited to select 10 patients with chronic diseases from the health files of community residents by simple random sampling and establish contact with patients by telephone, introduced the research content to them, and invited them to the community health service station to participate in this survey after obtaining their consent.

The participants included in the study met the following criteria: (1) had been living continuously in the rural community for more than 6 months, (2) were between 30 and 65 years old, (3) had been diagnosed with 1 or more chronic diseases, and (4) had experience using the internet. The exclusion criteria were as follows: (1) individuals with physical and mental conditions that made them not suitable for participation in this survey and (2) individuals who did not complete any 1 of the first 2 tasks.

Preparation for the Investigation

The surveys were conducted in person. Web-based survey software (Tongtai Questionnaire Survey Platform, Beijing Tongtai Technology Development Co Ltd) was used to complete the questionnaire survey, a small audiorecorder (model R2, JingZheng,) was used to record the interview content, and screen-recording software (Windows: KK Lu Xiang Ji, version 2.8; Android: Lu Ping Da Shi, version 3.3) was used to record the participants' behavior during the internet operational tasks. Before the start of the investigation, the screen-recording software was installed on the computer and tablet computer. After each participant completed the investigation, the investigator exported the recording file and cleared all browsing records and network settings. The duration of the investigation for each participant was limited to be within 1 hour. After each survey, we offered the participant a bucket of cooking oil worth 50 RMB (approximately US $7.74).

Statistical Analysis

Data entry and preprocessing were performed using Excel (version 2010; Microsoft Inc), and statistical analysis was performed using SPSS software (version 16.0, SPSS Inc). Descriptive statistical analyses were performed by calculating the mean and standard deviation, and intragroup differences

were analyzed using the *t* test and Kruskal-Wallis test. *P* values <.05 were considered significant.

Ethical Approval

This study was approved by the Ethics Committee of the Bengbu Medical College (2017054). The survey was completed anonymously. All potential respondents were contacted personally and thoroughly informed about the aim of the study, data processing, and the use of the data. Participation was voluntary, and participants could refuse to participate.

Results

Self-Rated Health Information Literacy

In total, 70 respondents (Table 1) completed task 1 and task 2, among whom 56 participants completed task 3, and 14 participants did not complete the semistructured interview (quit the interview due to loss of interest: n=7; did not give clear answers to the questions: n=5; provided irrelevant answers: n=2).

The average health information literacy score of the respondents was 46.21 (SD 4.89), out of a total possible score of 70; the highest score was 56, and the lowest score was 34. Of the 70 respondents, 91% (n=64) scored at least 40. The subdimension *health information evaluation* had the highest score (mean 14.30, SD 2.98), followed by *health information–seeking* (mean 12.80, SD 2.23), *health information consciousness* (mean 11.20, SD 1.04), and *health information application* (mean 7.91, SD 0.37).

The results of the univariate analysis showed that, except for gender (*P*=.04), education level (*P*=.02), and experience with internet use (*P*=.03), the self-rated health information literacy scores did not differ significantly based on the other sociodemographic factors (age: *P*=.70; occupation: *P*=.23; mobile phone use: *P*=.99; chronic diseases: *P*=.98; time postdiagnosis: *P*=.62).

Table 1. Sociodemographic characteristics and self-rated health information literacy scores.

Characteristic	Participants (n=70), n (%)	Self-rated health information literacy score	Chi-square or *t* test[a] (*df*)	*P* value
Gender			–2.135[a] (68)	.04
Male	39 (56)	45.13 (5.08)		
Female	31 (44)	47.58 (4.35)		
Age			0.156 (1)	.70
31 to 45 years old	27 (39)	46.44 (4.27)		
46 to 60 years old	33 (47)	46.33 (5.20)		
61 to 65 years old	10 (14)	45.20 (5.79)		
Education			7.799 (2)	.02
Primary school or below	9 (13)	42.78 (5.21)		
Middle school or technical school	56 (80)	47.09 (4.56)		
University	5 (7)	42.60 (4.39)		
Postgraduate and above	0 (0)	0.00 (0.00)		
Occupation			4.264 (3)	.23
Employees of enterprises and public institutions (including separation and retirement)	29 (41)	47.24 (4.33)		
Farming	10 (14)	44.30 (4.92)		
Commercial or service	19 (27)	45.16 (5.22)		
No job	12 (17)	47.00 (5.34)		
Experience with internet use			6.895 (2)	.03
Less than 1 year	12 (17)	43.00 (5.78)		
1-3 years	14 (20)	46.93 (4.68)		
More than 3 years	44 (63)	46.86 (4.45)		
Experience with mobile phone use			0.001(1)	.99
Less than 1 year	0 (0)	0.00 (0.00)		
1-3 years	6 (9)	45.83 (7.88)		
More than 3 years	64 (91)	46.25 (4.62)		
Number of chronic diseases			0.001 (1)	.98
1	42 (60)	46.17 (4.51)		
2 or more	28 (40)	46.29 (5.51)		
Time postdiagnosis			1.753 (3)	.62
Less than 1 year	7 (10)	45.00 (4.00)		
1-3 years	24 (34)	46.25 (6.04)		
4-5 years	12 (17)	46.25 (3.62)		
More than 5 years	27 (29)	46.92 (4.56)		

[a]A paired *t* test was used to obtain this value.

Behavioral Ability Test

Task 1

Of the 70 respondents, 64 (91%) were able to find websites containing health information on the internet, for which, 64% (41/64) chose to enter the URL of the health website directly. Another 36% of respondents (23/64) chose to search for health websites through search engines. Baidu (15/64, 23%) was the most popular search engine, and other search engines were chosen by 13% of respondents (8/64); 9% of respondents (6/70) did not complete this operational task. The primary reason for not completing this task was that they were not in the habit of using computers and mobile phones to search for information, and they just used WeChat for general social interaction.

Task 2

A total of 93% of the respondents (65/70) saved articles that they considered valuable. The content of the articles was mainly about disease treatment (73/185, 40%) and health care (54/185, 29%). The rest of the articles did not address health-related issues. However, 35% of the respondents (23/65) had difficulty saving articles because they did not know how to operate an internet browser.

Task 3

Similar to operational task 2, 93% of the respondents (65/70) completed this operational task. Among them, 88% (57/65) used search engines (with 30/57, 53% of those who used search engines using the Baidu search engine) to retrieve treatment-related articles for a certain chronic disease; other respondents searched for articles on certain health websites.

Out of a possible score of 75 points on the DISCERN scale, the mean score was 37.4 points (SD 13.8). Out of a possible score of 40 on part 1, the mean score was 23.9 (SD 9.3), and out of a possible score of 35 on part 1, the mean score was 13.6 (SD 6.0); 7 respondents scored below 10 points, 31 respondents scored between 20 and 40 points, and the other respondents scored between 40 and 60 points.

Semistructured Interviews

A total of 56 respondents completed semistructured interviews (Multimedia Appendix 2).

Internet Health Information Retrieval Strategy

Usage of Search Engines

The interview results showed that 96% of the interviewees (54/56) preferred to search for health information directly through search engines instead of choosing a special health website. This finding was consistent with the respondents' performance in operational task 3 and indicated that the awareness rate and selection rate of authoritative health websites were low. Some respondents (20/56, 36%) said the operational task was the first time they had used the internet to find articles about health, indicating that there are still some residents who do not actively use the internet to find health information and knowledge in rural China.

> *I feel Baidu is more credible in all websites, so need to find some information Baidu has inside.* [Participant 7, 45 years old]
>
> *Because the site involves a wide range of content, the feasibility, desirability, trust found in the viewing are better.* [Participant 39, 60 years old]
>
> *Because the site has the expert review mark, has the doctor basic information and has no advertisement.* [Participant 59, 56 years old]

Choice of Health Website

Although most respondents gave reasons why they chose a particular health information website, including the convenience of searching, personal habits, comprehensive content, and high reliability, 34% of the respondents (19/56) did not know why they chose the website and indicated that their website choice was random. We also assessed how the respondents determined whether a website was the best website to provide health information; 13% of respondents (7/56) believed that a website was the best if it contained a large amount of health information about diseases and treatments, but the respondents were not sure how to determine the authenticity of the articles on the website.

> *I also did not know which website to choose and had never visited this website before. In this investigation, I chose this website at random, so I could not judge whether this website was good or bad.* [Participant 9 years, 44 years old]
>
> *I have not used Baidu and other search engines, just in the search engine provided health sites, I also randomly selected, for how to evaluate the pros and cons of the site, I do not know.* [Participant 9, 44 years old]

Saving Internet Health Information

A few respondents (4/56, 7%) reported difficulties accessing health websites, which was similar to the findings of operational task 1. However, only 16% of respondents (9/56) thought that they had difficulties saving interesting articles, which was inconsistent with the findings of operational task 2.

> *It is not difficult to find the site, open a web browser to search for such articles can be easily found. It was not difficult to save articles before, but now I find that some articles need to register and log in when they are saved, or even charge for it, which makes me feel more cumbersome.* [Participant 22, 51 years old]

Evaluation of the Quality of Internet Health Information

Layout of the Website

Overall, 75% of the respondents (42/56) thought the layout of the web page was well designed, and 55% of the respondents (31/56) were satisfied with the website in terms of a detailed and clear layout, clear classification, comprehensive search information, provision of answers internet by professional doctors, and a lack of advertising plug-ins and links; however, 32% of the respondents (18/56) were quite dissatisfied with the large number of advertising links interspersed in the web pages.

> *I feel that the layout design of this kind of health information website is similar. What I am most dissatisfied with is that when I browse the web page, advertisements for doctor consultation often pop up, and they will pop up again and again after I click "close."* [Participant 2, 32 years old]
>
> *I think the homepage of the website is not bad, and the classification of diseases is obvious. What I am not satisfied with is that the font of the page is relatively small and there are many advertisements, which affect the professionalism.* [Participant 61, 47 years old]

Value of Internet Health Information Content

Of the 56 respondents, 96% (54/56) said they had acquired new knowledge from health information websites, and 6% (36/56)

believed that information about disease treatment, rehabilitation, health care and prevention was the most valuable information to them. Additionally, 91% (51/56) said they would visit the website again and recommend it to their relatives and friends.

> *The most helpful information on this website is the knowledge of disease prevention and improvement. I also learned a lot of knowledge that I didn't know before from the internet. It is very likely that I will visit this website again, or will I recommend it to my family.* [Participant 30, 40 years old]

Evaluation of the Quality of Internet Health Information

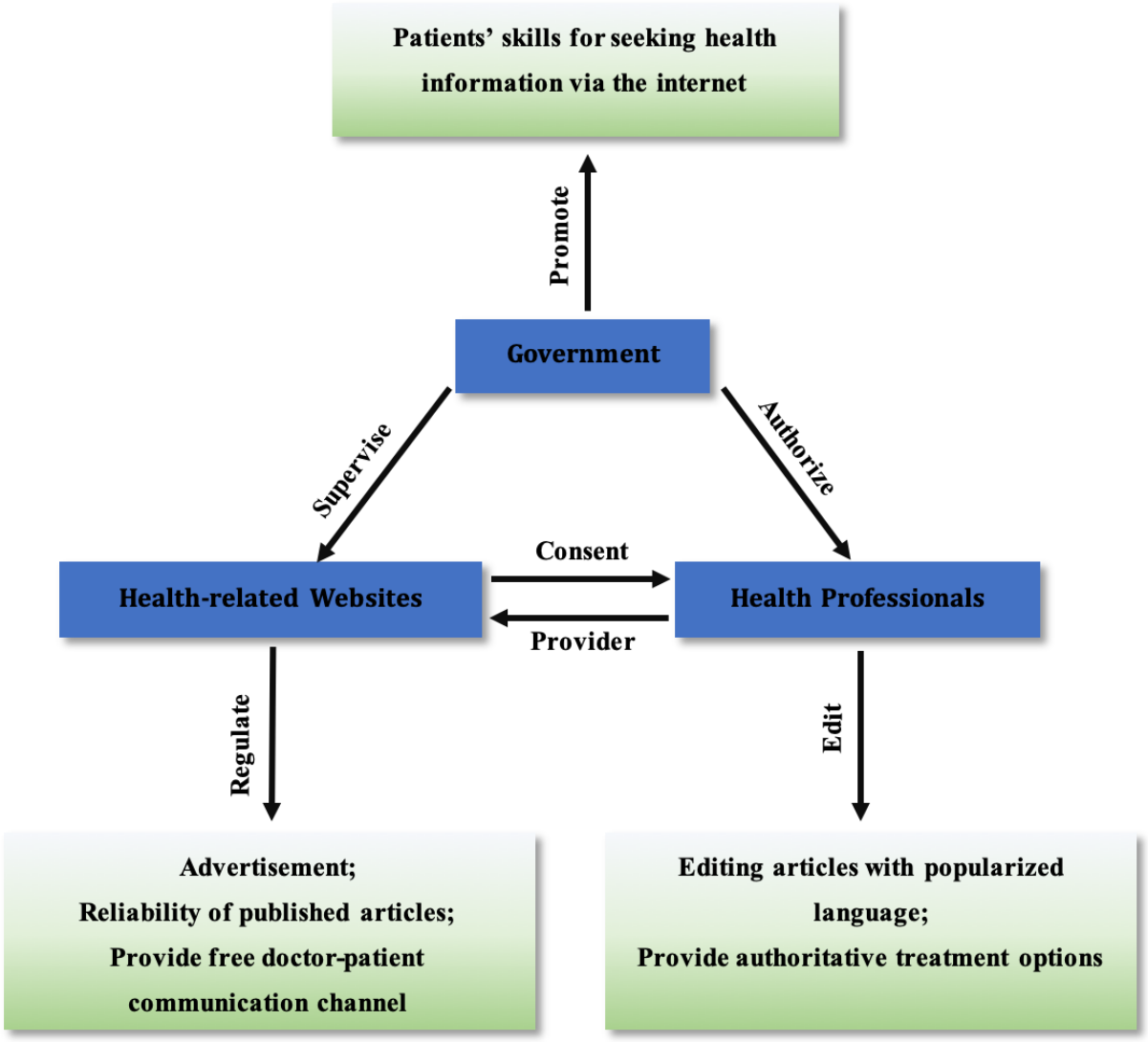
A total of 89% of respondents (50/56) were able to use some method to determine the reliability of the information, while 11% (6/56) said they did not know how to judge the reliability of the various articles on the website.

> *First of all, the website has the sign of expert review. Secondly, if it is not a regular website, there will be a variety of advertising push, while regular health websites generally have no advertising push.* [Participant 59, 56 years old]

Comments and Suggestions

At the end of the interview, we conducted an analysis of the interview recordings and summarized the interviewees' opinions and suggestions on obtaining internet health information (Figure 1).

Figure 1. Interviewee suggestions.



Discussion

Principal Results

In this study, qualitative and quantitative research methods were used to investigate the status quo of access to internet health information among 70 rural community patients with chronic diseases to comprehensively assess their health information literacy levels and ability to access internet health information and to summarize the obstacles they faced in accessing internet health information.

The self-rated health information literacy score of all the interviewees was above 30 points, with more than 90% of respondents (63/70) scoring above 40, indicating that most respondents had a moderate health information literacy level.

The health information literacy levels of the respondents over 60 years old were lower than those of younger respondents, and there were significant differences in the health information literacy scores of the respondents of different sexes, education levels and internet use experiences, which is similar to the findings of previous studies [18,20,21]. Health information literacy is an important aspect of individual cultural literacy, and there is an interactive relationship between health information attitude and health information skills [27]. Therefore, an important aspect of future health education and health promotion is enhancing the awareness of patients with chronic disease on the value of health information and cultivating their practical ability to obtain internet health information.

The results of operational task 1 and operational task 2 indicated that most respondents had a low ability to access health information. The respondents randomly chose health websites and could not judge whether the website was reliable based on the number of visits to the website and the authority of the articles. Most articles selected by the respondents came from websites with encyclopedic knowledge or answers from people based on their own experiences rather than authoritative health information websites. The respondents' abilities to download or save interesting articles were poor. One important reason for this finding is that people's use of the internet is mostly for entertainment and leisure (such as using Tiktok and WeChat [28]). This only requires people to know the basics of Android or iOS (such as opening an app, returning to the home page, and selecting and typing text). People use their mobile phones as a tool to query internet health information only when health problems occur [29], and proficiency of this tool requires the support of certain operational abilities (such as using the browser, opening a web page, downloading files, saving files, setting up software).

In terms of the quality of the internet health information obtained, most health-related articles saved by the respondents did not specify the author, publisher, publication date, or references. The articles obtained via search engines were primarily written based on people's experiences, and most treatment protocols involved were based on the subjective opinions of individuals and failed to describe the effects and risks of the treatment methods. Several articles published on websites not related to medical health were even chosen as describing the best treatment options. Although a large number of internet evaluation tools have been developed, these tools are mostly designed from the perspective of expert evaluation, and there are few public-facing evaluation systems [30]. Therefore, it is necessary to develop public-facing internet health information quality evaluation tools as soon as possible. Additionally, the results of the DISCERN evaluation indicated that most respondents lacked the skills to access high-quality health-related information via the internet. Combining the results of the health literacy self-assessment with those of the DISCERN evaluation, we found that most respondents overrated their ability to seek high-quality health information internet.

In further conversations with 56 of the respondents, we found that most had directly retrieved health-related information through search engines (such as Baidu), while the proportion of health-related information obtained from professional health sites and medical sites was relatively low. Most interviewees believed that they could retrieve professional health information from the websites. The popularity of the websites, the quality of the articles, the illness of the interviewees and the advice of their doctors were the main reasons for their choice of internet health information. While 16% of interviewees (9/56) admitted that they had difficulties saving and downloading health information, this proportion reached 35% (23/65) in operational task 2, which contradicted the respondents' self-assessments. The respondents' evaluation of internet health information quality relied more on their own judgements than on other factors because they did not compare health information from multiple channels and seldom consulted professionals. Most respondents rated health-related websites based on the reasonableness of the page layout and whether there were spam ads. They were more concerned with the richness and comprehensibility of the website content than the authenticity and reliability of the content. This finding indicates that there is a need to strengthen education on the awareness for internet health information evaluation among patients with chronic diseases. Simultaneously, as providers of internet health information, health websites should also strengthen quality control of their health information and improve the publicity of their website so that consumers can more easily find the high-quality internet health information that they need.

Limitations

This study had some limitations. First, this study investigated only 70 patients with chronic disease in 10 communities of a small town near Bengbu City. The sample size was relatively small, and the scope of the research was limited, which limits the generalizability of the results to some extent. Second, this study designed only 3 tasks to evaluate the residents' behavioral ability to obtain internet health information, which limits the ability to fully reflect the actual level of rural residents' behavioral abilities to obtain internet health information. Finally, some bias was inevitable due to the influence of differences in social and cultural backgrounds between interviewers and interviewees who understand the questions differently during the semistructured interview process. Therefore, the results of this study should be regarded as preliminary and interpreted with caution. Despite these limitations, the findings of this study are valuable for understanding the internet health information–seeking behavior of patients with chronic diseases in rural China and their attitudes about internet health information.

Practical Implications

Governments and relevant departments should strengthen education for the general public to help them increase their ability to access internet health information and access free and high-quality internet health information resources. Governments and relevant departments should also increase supervision on the release and dissemination of internet health information with relevant laws and regulations to improve the internet health information environment.

Conclusion

We found that the level of internet information access among patients with chronic disease living in rural China was moderate. The vast majority of the respondents recognized the importance of accessing health information but were not very proactive in accessing health information via the internet. Furthermore, there was a gap between their actual ability to access high-quality internet health information and their self-rated health information literacy. Most respondents experienced difficulties seeking internet health information, and they lacked the skills to screen for high-quality internet information. Although most interviewees listed certain methods for judging the quality of internet health information, the behavioral ability test showed that they did not follow the expected methods to obtain high-quality internet health information but rather only made subjective judgements and showed a certain degree of randomness in their selection processes.

Acknowledgments

This study was supported by funding from the National Planning Office of Philosophy and Social Sciences of the People's Republic of China (grants 17BGL262 and 17AZD037) and the Talent Cultivation Program of Bengbu Medical College (51201209). We thank American Journal Experts for language editing.

Authors' Contributions

The study was conceived and designed by FW and ZW. The cross-sectional surveys and interviews were conducted by ZW, YF, HL, and SD. HX and LZ drafted and revised the manuscript for important intellectual content, and FW and AL approved the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Investigation materials for patients with chronic diseases.
[DOCX File , 21 KB - jmir_v24i1e26308_app1.docx]

Multimedia Appendix 2
Analysis of the semistructured interview results of the 56 interviewees.
[DOCX File , 19 KB - jmir_v24i1e26308_app2.docx]

References

1. Hunter DJ, Reddy KS. Noncommunicable diseases. N Engl J Med 2013 Oct 03;369(14):1336-1343. [doi: 10.1056/NEJMra1109345] [Medline: 24088093]
2. Noncommunicable diseases. World Health Organization. URL: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases [accessed 2020-11-05]
3. Lai S, Gao J, Zhou Z, Yang X, Xu Y, Zhou Z, et al. Prevalences and trends of chronic diseases in Shaanxi Province, China: evidence from representative cross-sectional surveys in 2003, 2008 and 2013. PLoS One 2018 Aug 23;13(8):e0202886 [FREE Full text] [doi: 10.1371/journal.pone.0202886] [Medline: 30138444]
4. Ebele NA, Obiora CN. Extent of access to health information and sources for chronic disease patients in tertiary health institutions in south east Nigeria: implications for libraries role. Libr Philos Pract 2017 Jun:1-16 [FREE Full text]
5. LaValley SA, Kiviniemi MT, Gage-Bouchard EA. Where people look for online health information. Health Info Libr J 2017 Jun;34(2):146-155 [FREE Full text] [doi: 10.1111/hir.12143] [Medline: 27207817]
6. James DCS, Harville C. eHealth literacy, online help-seeking behavior, and willingness to participate in mhealth chronic disease research among African Americans, Florida, 2014-2015. Prev Chronic Dis 2016 Nov 17;13:E156 [FREE Full text] [doi: 10.5888/pcd13.160210] [Medline: 27854421]
7. Qiu Y, Ren W, Liu Y, Yin P, Ren J. Online health information in a rural residential population in Zhejiang Province, China: a cross-sectional study. BMJ Open 2019 May 05;9(5):e026202 [FREE Full text] [doi: 10.1136/bmjopen-2018-026202] [Medline: 31061032]
8. Rogers MA, Lemmen K, Kramer R, Mann J, Chopra V. Internet-delivered health interventions that work: systematic review of meta-analyses and evaluation of website availability. J Med Internet Res 2017 Mar 24;19(3):e90 [FREE Full text] [doi: 10.2196/jmir.7111] [Medline: 28341617]
9. Wang F, Wang Z, Sun W, Yang X, Bian Z, Shen L, et al. Evaluating the quality of health-related WeChat public accounts: cross-sectional study. JMIR Mhealth Uhealth 2020 May 08;8(5):e14826 [FREE Full text] [doi: 10.2196/14826] [Medline: 32383684]

10. Kuenzel U, Monga Sindeu T, Schroth S, Huebner J, Herth N. Evaluation of the quality of online information for patients with rare cancers: thyroid cancer. J Cancer Educ 2018 Oct;33(5):960-966. [doi: 10.1007/s13187-017-1173-z] [Medline: 28120139]
11. Cisu TI, Mingin GC, Baskin LS. An evaluation of the readability, quality, and accuracy of online health information regarding the treatment of hypospadias. J Pediatr Urol 2019 Feb;15(1):40.e1-40.e6. [doi: 10.1016/j.jpurol.2018.08.020] [Medline: 30449679]
12. Fuzhi W, Dan L, Weiwei S, Tingting Y, Dehua H, Wei P, et al. Health information literacy and barriers of online health information seeking among digital immigrants in Rural China: a preliminary survey. SAGE Open 2019 Jun 12;9(2):215824401985694. [doi: 10.1177/2158244019856946]
13. Zhang S, Du J. Health information literacy should become the critical point of health promotion for Chinese public. Journal of Medical Informatics 2010;31(02):45-49.
14. Wang F, Luo A, Xie W. Connotation of public health information literacy in China and its training strategies. Chin Journal Medical Library Information 2013;22(08):13-17.
15. Schardt C. Health information literacy meets evidence-based practice. J Med Libr Assoc 2011 Jan;99(1):1-2 [FREE Full text] [doi: 10.3163/1536-5050.99.1.001] [Medline: 21243047]
16. Gutierrez N, Kindratt TB, Pagels P, Foster B, Gimpel NE. Health literacy, health information seeking behaviors and internet use among patients attending a private and public clinic in the same geographic area. J Community Health 2014 Feb;39(1):83-89. [doi: 10.1007/s10900-013-9742-5] [Medline: 23900880]
17. Ghweeba M, Lindenmeyer A, Shishi S, Abbas M, Waheed A, Amer S. What predicts online health information-seeking behavior among Egyptian adults? a cross-sectional study. J Med Internet Res 2017 Jun 22;19(6):e216 [FREE Full text] [doi: 10.2196/jmir.6855] [Medline: 28642216]
18. Tan SS, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017 Jan 19;19(1):e9 [FREE Full text] [doi: 10.2196/jmir.5729] [Medline: 28104579]
19. Din HN, McDaniels-Davidson C, Nodora J, Madanat H. Profiles of a health information-seeking population and the current digital divide: cross-sectional analysis of the 2015-2016 California health interview survey. J Med Internet Res 2019 May 14;21(5):e11931 [FREE Full text] [doi: 10.2196/11931] [Medline: 31094350]
20. Sedrak MS, Soto-Perez-De-Celis E, Nelson RA, Liu J, Waring ME, Lane DS, et al. Online health information-seeking among older women with chronic illness: analysis of the women's health initiative. J Med Internet Res 2020 Apr 09;22(4):e15906 [FREE Full text] [doi: 10.2196/15906] [Medline: 32271152]
21. Rooks RN, Kapral CG, Mathis AL. Chronic conditions may be more important than race or ethnicity in relation to health information seeking and use. J Aging Health 2019 Apr;31(4):611-630. [doi: 10.1177/0898264317744643] [Medline: 29254425]
22. Sun Q, Zhou W, Zhang Y, Zhang F. Investigation and analysis of health information-seeking behavior among patients with chronic disease. Journal of Nursing Science 2019 May 5;34(9):84-86.
23. Zhoujie, Lixiaoping, Hudehua. Characteristics and Influence Factors of Health Information Seeking Behaviors of Hyperpietics in Minority Areas. Journal Of Medical Informatics 2019 Jul 25:73-76+85.
24. Eriksson-Backa K, Ek S, Niemelä R, Huotari M. Health information literacy in everyday life: a study of Finns aged 65-79 years. Health Informatics J 2012 Jun;18(2):83-94 [FREE Full text] [doi: 10.1177/1460458212445797] [Medline: 22733677]
25. Luo D, Zhou H, Yang X, Sun W, Wang F. Modification and primary application of Everyday Health Information Literacy (EHIL-10) Self-rate Questionnaire. J Baotou Medical College 2019 Sep 15;35(9):81-85+88.
26. Charnock D, Shepperd S, Needham G, Gann R. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices. J Epidemiol Community Health 1999 Feb;53(2):105-111 [FREE Full text] [Medline: 10396471]
27. Wang F, Luo A, Luo D, Hu D, Xie W, Chen Z, et al. The association between health information attitudes and skills in patients with chronic disease in China. Glob Health Promot 2016 Dec 09;23(4):59-69. [doi: 10.1177/1757975915577922] [Medline: 26041814]
28. Shen L, Wang S, Chen W, Fu Q, Evans R, Lan F, et al. Understanding the function constitution and influence factors on communication for the WeChat official account of top tertiary hospitals in China: cross-sectional study. J Med Internet Res 2019 Dec 09;21(12):e13025 [FREE Full text] [doi: 10.2196/13025] [Medline: 31815674]
29. Thapa DK, Visentin DC, Kornhaber R, West S, Cleary M. The influence of online health information on health decisions: a systematic review. Patient Educ Couns 2021 Apr;104(4):770-784. [doi: 10.1016/j.pec.2020.11.016] [Medline: 33358253]
30. Sun Y, Zhang Y, Gwizdka J, Trace CB. Consumer evaluation of the quality of online health information: systematic literature review of relevant criteria and indicators. J Med Internet Res 2019 May 02;21(5):e12522 [FREE Full text] [doi: 10.2196/12522] [Medline: 31045507]

Abbreviations

RMB: Renminbi

Edited by G Eysenbach; submitted 06.12.20; peer-reviewed by W Sun, L Shen, P Yin, H Imeri; comments to author 14.01.21; revised version received 28.02.21; accepted 28.12.21; published 31.01.22.

Please cite as:
Wang Z, Fan Y, Lv H, Deng S, Xie H, Zhang L, Luo A, Wang F
The Gap Between Self-Rated Health Information Literacy and Internet Health Information-Seeking Ability for Patients With Chronic Diseases in Rural Communities: Cross-sectional Study
J Med Internet Res 2022;24(1):e26308
URL: https://www.jmir.org/2022/1/e26308
doi:10.2196/26308
PMID:35099401

Original Paper

Energy Efficiency of Inference Algorithms for Clinical Laboratory Data Sets: Green Artificial Intelligence Study

Jia-Ruei Yu[1], MD; Chun-Hsien Chen[1,2], PhD; Tsung-Wei Huang[3], PhD; Jang-Jih Lu[1], MD, PhD; Chia-Ru Chung[4], PhD; Ting-Wei Lin[1], MD; Min-Hsien Wu[5], PhD; Yi-Ju Tseng[6*], PhD; Hsin-Yao Wang[1,5*], MD

[1]Department of Laboratory Medicine, Chang Gung Memorial Hospital at Linkou, Taoyuan City, Taiwan

[2]Department of Information Management, Chang Gung University, Taoyuan City, Taiwan

[3]Department of Electrical and Computer Engineering, University of Utah, Salt Lake City, UT, United States

[4]Department of Computer Science and Information Engineering, National Central University, Taoyuan City, Taiwan

[5]Graduate Institute of Biomedical Engineering, Chang Gung University, Taoyuan City, Taiwan

[6]Department of Information Management, National Central University, Taoyuan City, Taiwan

[*]these authors contributed equally

Corresponding Author:
Hsin-Yao Wang, MD
Department of Laboratory Medicine
Chang Gung Memorial Hospital at Linkou
No 5 Fuxing Street
Guishan District
Taoyuan City, 333
Taiwan
Phone: 886 978112962
Fax: 886 33281200
Email: mdhsinyaowang@gmail.com

Abstract

Background: The use of artificial intelligence (AI) in the medical domain has attracted considerable research interest. Inference applications in the medical domain require energy-efficient AI models. In contrast to other types of data in visual AI, data from medical laboratories usually comprise features with strong signals. Numerous energy optimization techniques have been developed to relieve the burden on the hardware required to deploy a complex learning model. However, the energy efficiency levels of different AI models used for medical applications have not been studied.

Objective: The aim of this study was to explore and compare the energy efficiency levels of commonly used machine learning algorithms—logistic regression (LR), k-nearest neighbor, support vector machine, random forest (RF), and extreme gradient boosting (XGB) algorithms, as well as four different variants of neural network (NN) algorithms—when applied to clinical laboratory datasets.

Methods: We applied the aforementioned algorithms to two distinct clinical laboratory data sets: a mass spectrometry data set regarding *Staphylococcus aureus* for predicting methicillin resistance (3338 cases; 268 features) and a urinalysis data set for predicting *Trichomonas vaginalis* infection (839,164 cases; 9 features). We compared the performance of the nine inference algorithms in terms of accuracy, area under the receiver operating characteristic curve (AUROC), time consumption, and power consumption. The time and power consumption levels were determined using performance counter data from Intel Power Gadget 3.5.

Results: The experimental results indicated that the RF and XGB algorithms achieved the two highest AUROC values for both data sets (84.7% and 83.9%, respectively, for the mass spectrometry data set; 91.1% and 91.4%, respectively, for the urinalysis data set). The XGB and LR algorithms exhibited the shortest inference time for both data sets (0.47 milliseconds for both in the mass spectrometry data set; 0.39 and 0.47 milliseconds, respectively, for the urinalysis data set). Compared with the RF algorithm, the XGB and LR algorithms exhibited a 45% and 53%-60% reduction in inference time for the mass spectrometry and urinalysis data sets, respectively. In terms of energy efficiency, the XGB algorithm exhibited the lowest power consumption for the mass spectrometry data set (9.42 Watts) and the LR algorithm exhibited the lowest power consumption for the urinalysis data set (9.98 Watts). Compared with a five-hidden-layer NN, the XGB and LR algorithms achieved 16%-24% and 9%-13% lower power

consumption levels for the mass spectrometry and urinalysis data sets, respectively. In all experiments, the XGB algorithm exhibited the best performance in terms of accuracy, run time, and energy efficiency.

Conclusions: The XGB algorithm achieved balanced performance levels in terms of AUROC, run time, and energy efficiency for the two clinical laboratory data sets. Considering the energy constraints in real-world scenarios, the XGB algorithm is ideal for medical AI applications.

(J Med Internet Res 2022;24(1):e28036) doi:10.2196/28036

KEYWORDS

medical informatics; machine learning; algorithms; energy consumption; artificial intelligence; energy efficient; medical domain; medical data sets; informatics

Introduction

Machine learning (ML) methods have been successfully employed in various medical fields [1-5], and energy consumption during ML inference has been attracting increasing attention [6-8]. The increasing focus on inference energy can primarily be attributed to two reasons. First, energy constraints constitute a major issue when ML is deployed into battery-powered medical devices [9-11]. Second, to achieve high predictive performance, the computation and memory requirements of ML models have increased. The growth of model size has been well reflected in neural networks (NNs) over the last decade, which are considered as the main ML algorithms implemented during this period.

An optimal ML model should achieve balanced predictive performance and energy efficiency. However, most relevant studies have only focused on comparing the predictive performance of different ML algorithms [12-14] and have not thoroughly explored the energy efficiency of different ML algorithms in the medical domain. Data formats in the medical field are diverse, and clinical laboratory data are a common type of medical data. In real-world settings, single laboratory tests must be subjected to strict validation procedures before their clinical use. Thus, the data obtained from such tests usually comprise features that are highly associated with the prediction targets. The characteristics of clinical laboratory data sets are unique, and the energy efficiency of different ML algorithms for processing clinical laboratory data sets warrants investigation.

A partial explanation for the poor understanding of energy efficiency is that estimating energy consumption is more difficult than estimating other metrics (eg, accuracy) [15]. Several methods exist for evaluating the energy consumption of ML models. Computational complexity can be used for theoretically approximating the number of operations; thus, it can be used to estimate energy consumption (Multimedia Appendix 1) [16-18]. Studies have established formulas for estimating energy consumption; these formulas sum the energy consumption levels of different elementary operations on the basis of complexity theory and benchmark results [7,19]. However, these formulas are available for only specific ML models and cannot be expanded to all algorithms. In addition to the aforementioned estimation formulas, experimental approaches can be used for estimating energy consumption. Currently, simulation and performance counters are the two main approaches for experimentally estimating energy consumption [15]. Although simulations enable fine-grained energy estimation at the architecture and instruction levels, the use of simulations for large-scale ML tasks is not feasible due to the considerable overhead involved [15]. By contrast, performance counters, which are a set of registers in processors that log specific hardware-related events, do not generate any overhead; therefore, these counters are suitable for use in different ML applications.

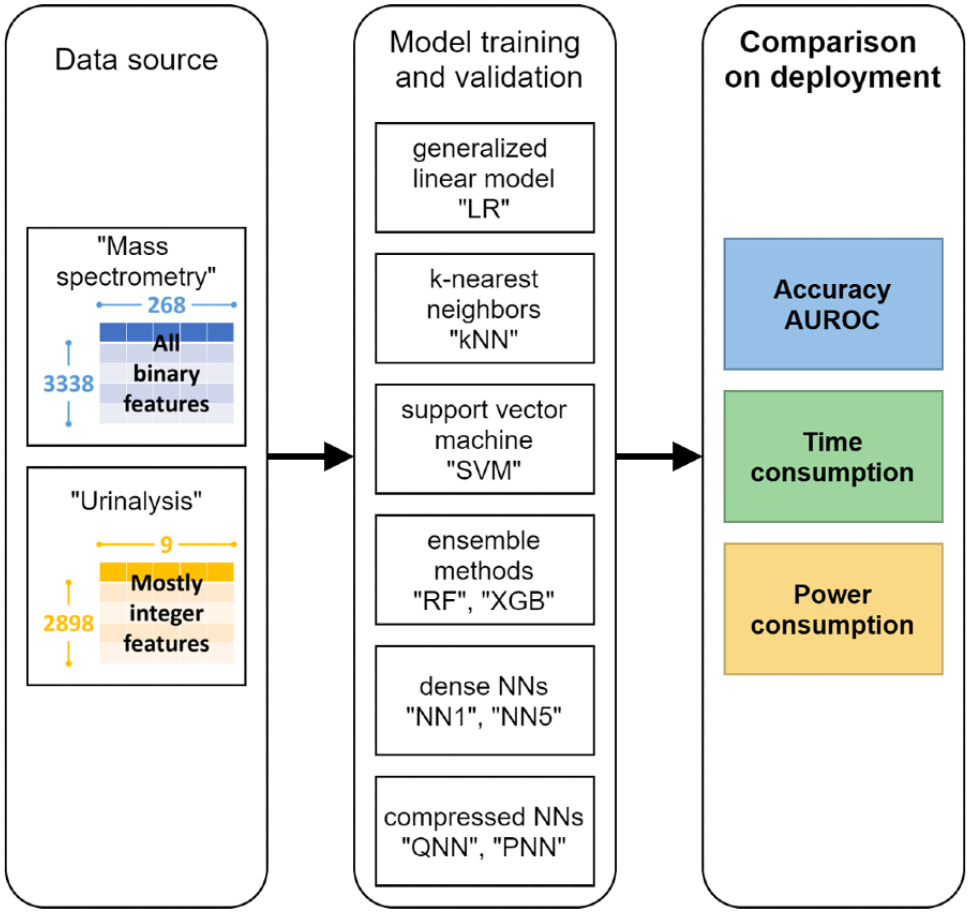
In this study, we estimated the power consumption of nine algorithms during ML inference: logistic regression (LR), k-nearest neighbor (kNN), support vector machine (SVM), random forest (RF), extreme gradient boosting (XGB), and four NN-based algorithms. These algorithms were used to classify two clinical laboratory data sets: a large binary feature set and a small integer feature set. The following performance measures were recorded: accuracy, area under the receiver operating characteristic curve (AUROC), time consumption, and power consumption. The time and power consumption were determined using performance counter data from Intel Power Gadget 3.5. Finally, we performed statistical tests to validate our results. The results indicated the energy efficiency of each investigated ML algorithm in medical applications.

Methods

Study Design and Environmental Settings

Figure 1 illustrates the process flowchart of this study. We used two preprocessed data sets to train ML models with the nine considered algorithms: a mass spectrometry data set, based on matrix-assisted laser desorption/ionization time-of-flight mass spectrometry data of *Staphylococcus aureus* for predicting methicillin resistance, and a urinalysis data set, based on urinalysis data for predicting *Trichomonas vaginalis* infection. Subsequently, we comprehensively evaluated the trained models in terms of predictive performance, time consumption, and power consumption using independent testing data.

Figure 1. Process flowchart of this study. LR: logistic regression; kNN: k-nearest neighbor; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting; NN1: one-hidden-layer neural network; QNN: quantized five-hidden-layer neural network; PNN: pruned five-hidden-layer neural network; NN5: five-hidden-layer neural network; AUROC: area under the receiver operating characteristic curve.



All experiments were run on a Windows 10 personal computer with 4 GB RAM and a 2.3 GHz Intel Core i5-8300H central processing unit (CPU). All ML models were implemented using Python 3.7.1 with the following Python libraries: scikit-learn 0.23.1 [20], xgboost 0.90 [21], and pytorch 1.8.1 [22]. Intel Power Gadget 3.5 [23] was used to acquire time and power measurements. Additional details regarding the power measurement and Intel Power Gadget 3.5 are provided in the "Time Consumption and Power Consumption" section below. All statistical analyses were performed using the "rstatix" package of R software (version 4.0.2).

Data Source

Data Set Characteristics

The mass spectrometry and urinalysis data sets adopted in this study represent distinct feature patterns in ML; Table 1 presents their characteristics. The mass spectrometry data set has a relatively large feature set comprising 268 binary features. By contrast, the urinalysis data set has a small feature set comprising only nine features, almost all of which are integer features in a larger range. These data sets have been applied and validated in previous studies [24-26].

Table 1. Characteristics of the final mass spectrometry and urinalysis data sets.

Data set	Cases, n	Features, n	Binary features, n	Integer features, n	Majority class	Percentage of majority class (Nmax/N)	Gini Impurity
Mass spectrometry	3338	268	268	0	Methicillin-resistant *Staphylococcus aureus*	53.0%	0.50
Urinalysis	2898	9	2	7	*Trichomonas vaginalis*-negative	57.1%	0.49

Mass Spectrometry Data Set

The mass spectrometry data set comprises mass spectral information on methicillin-resistant and methicillin-sensitive *S. aureus* isolates. We collected routine mass spectrometry data about *S. aureus* samples consecutively from Chang Gung Memorial Hospital in 2016, and we identified the methicillin

resistance of every *S. aureus* isolate by employing the paper disk method with cefoxitin.

In the original mass spectral data, intensity values are a function of the mass-to-charge ratio. We preprocessed these data using a validated binning method [24]. After data preprocessing, every feature in the mass spectrometry data set was determined to correspond to a 10-Da interval of mass-to-charge ratio. All the features are binary features, and the values 1 and 0 represent the presence and absence of a peak, respectively (ie, the presence and absence of a sufficient intensity, respectively), in a specific interval. The final size of the mass spectrometry data set was 3338×268 entries, with no missing values.

Urinalysis Data Set

The urinalysis data set comprises routine urinalysis data (including leukocyte esterase, nitrite, protein, occult blood, red blood cell count, white blood cell count, and epithelial cell count) and demographic data (age and sex) of patients with and without *T. vaginalis* infections. The diagnosis of *T. vaginalis* infection was made according to microscopic tests. The urinalysis data set comprises the data of all patients who received at least one urinalysis test at Chang Gung Memorial Hospital between January 2009 and December 2013. The original data set consists of 839,164 cases; because the outcome distribution is imbalanced in the original data [26], we applied random undersampling and the synthetic minority oversampling technique [27]. The final size of the urinalysis data set was 2898×9 entries and the percentage of the majority class was 57.1%.

In the urinalysis data set, “sex” and “nitrite” are binary features, which are represented by 1 and 0. “Leukocyte esterase,” “protein,” and “occult blood” data are semiquantitative features, which are represented on scales ranging from 0 to 4 (“negative,” “trace,” 1+, 2+, and 3+), 0 to 5 (“negative,” “trace,” 1+, 2+, 3+, and 4+), and 0 to 5 (“negative,” “trace,” 1+, 2+, 3+, and 4+), respectively. “Age,” “red blood cell count,” “white blood cell count,” and “epithelial cell count” are nonnegative integer features with maximum values of 103, 501, 501, and 101, respectively. The urinalysis data set does not contain missing values, and we did not perform further feature selection for this data set.

Model Training and Validation

Algorithms

LR Algorithm

LR is one of the simplest binary classification algorithms. In the LR algorithm, the predictive outcome $\hat{y}(\mathbf{x})$ of given data $\mathbf{x}$ is defined as follows:

$$\hat{y}(x)=[1+\exp(w_0+\mathbf{w}^T\mathbf{x})]^{-1}$$

where $\mathbf{w}$ and $\mathbf{w_0}$ represent the weight vector and bias of the LR model, respectively. LR is an example of a generalized linear model, and the output of the LR model represents the estimated probability of a certain class [17].

kNN Algorithm

The kNN algorithm is a memory-based algorithm. Accordingly, predictions of the kNN algorithm are directly based on the training data set and no additional training is required [28]. Predictions of the kNN algorithm, which are denoted by $\hat{y}(\mathbf{x})$, are based on the voting results of the k most similar instances in the training dataset [29]. The parameter $\hat{y}(\mathbf{x})$ is defined as follows:

where K denotes the set of k instances in the training data set that are most similar to the given data $\mathbf{x}$ and $\mathbf{w}(\mathbf{x_i})$ denotes the weighted value of the corresponding $\mathbf{x}'$ value. In the kNN algorithm, the distance between two data points $\mathbf{x}$ and $\mathbf{x}'$ is typically defined as follows:

where q is a given positive number. The Manhattan distance (for q=1) and the Euclidean distance (for q=2) are the two most common distance metrics. In this study, the numbers of nearest neighbors (k) were 27 and 7 in the final models of the mass spectrometry and urinalysis data sets, respectively.

SVM Algorithm

The SVM algorithm is a commonly used binary classification method. The purpose of the SVM algorithm is to find a hyperplane that separates two classes of data with the maximum margin in the feature space [29]. In the original linear SVM, the output binary features are labeled as +1 and −1, and the predictive outcome $\hat{y}(\mathbf{x})$ of given data $\mathbf{x}$ is defined as follows:

$$\hat{y}(\mathbf{x})=\text{sign}(w_0+\mathbf{w}^T\mathbf{x})$$

where $\mathbf{w}$ and w_0 represent the weight vector and bias of the SVM model, respectively.

The SVM algorithm is frequently applied with kernel transformations. Kernel functions represent the similarity between two data points. The radial basis function is one of the most commonly used kernel functions and is defined as follows:

$$\kappa(\mathbf{x},\mathbf{x}')=\exp(-\|\mathbf{x}-\mathbf{x}'\|^2/2\sigma^2)$$

where σ is the bandwidth. Through kernel transformations, the original feature space can be mapped into a higher dimension, which may improve the predictive performance of the SVM algorithm. For a kernelized SVM algorithm, the following equation is obtained:

where α is a sparse vector. For all nonzero α_i values, corresponding $\mathbf{x_i}$ terms represent support vectors. Accordingly, the final prediction $\hat{y}(\mathbf{x})$ depends on only the support vectors and is independent of the remaining training data. In this study, we selected the kernelized SVM algorithm with the radial basis function kernel as our final SVM model according to its validation performance.

RF Algorithm

RF is an ensemble decision tree classifier. The prediction of the RF algorithm, namely $\hat{y}(\mathbf{x})$, depends on the voting results of numerous decision trees [12]. The parameter $\hat{y}(\mathbf{x})$ is defined as follows:

where T represents the set of decision trees in the RF and $T_i(\mathbf{x})$ represents the predictive outcome of a given decision tree.

RF training involves the "bagging" (ie, bootstrap aggregating) technique [30]. Accordingly, each decision tree in the RF considers only a subset of training cases to improve model generalizability. In addition to the bagging technique, each split of the decision trees only considers a subset of the input features during training to prevent the growth of highly correlated trees [31]. In this study, the number of trees was set to 1000 and 1500 in the final RF models for the mass spectrometry and urinalysis data sets, respectively.

XGB Algorithm

The XGB algorithm is a type of ensemble algorithm, which uses the "boosting" technique to reduce the overall bias by sequentially combining weak classifiers into a model [32]. In practice, shallow decision trees are typical weak classifiers. The outcome $\hat{y}(\mathbf{x})$ of XGB models represents the log odds ratio of a certain class in binary classification tasks. The parameter $\hat{y}(\mathbf{x})$ is defined as follows:

where $T_i(\mathbf{x})$ denotes the *M* decision tree regressors in the model.

In each iteration, the training of decision trees in XGB is equivalent to a process of minimizing a certain objective function. Because XGB is a regularized algorithm [21], its objective function is different from those of the original gradient boosting algorithms. The objective function of a given decision tree in the XGB algorithm can be expressed as follows:

where N denotes the number of leaf nodes in a decision tree and γ and λ denote given positive numbers. The parameters G_j and H_j are defined as follows:

where $\hat{y}_i$ represents the predictive outcome of training data $\mathbf{x_i}$ after a certain number of iterations, $l(y_i,\hat{y}_i)$ represents a certain loss function between the predictive and actual outcomes, and L_j represents the set of data points $\mathbf{x}_i$ belonging to the *j*th leaf node.

In this study, we implemented the XGB model by using the "xgboost 0.90" Python library.

NN Algorithms

NN is a type of ML model that is inspired by the human nervous system. An NN consists of multiple layers of nodes (or neurons). The layer that receives the initial data is the input layer and the layer that exports the predictive results is the output layer. Numerous hidden layers exist between the input and output layers. The outputs of each node in an NN are obtained according to the outputs of the nodes in the previous layer [12].

Several types of connection patterns are possible between two adjacent layers. For example, in a classic fully connected layer, a series of weighted sums of the inputs is first calculated according to the given model parameters. These weighted sums are subjected to nonlinear transformation to obtain the output of the aforementioned layer. In practice, these steps are implemented using vectorized expressions, and the output vector of the *n*th hidden layer, namely $\mathbf{a_{n+1}}$, is expressed as follows:

$$\mathbf{a_{n+1}}=g(\Theta_n \cdot \mathbf{a_n}+\mathbf{b_n})$$

where $\mathbf{a_n}$, Θ_n, and $\mathbf{b_n}$ represent the input vector, weight matrix, and bias vector of the hidden layer, respectively, and g represents a nonlinear activation function (eg, the sigmoid function or rectified linear unit activation function).

NNs are ML models that are flexible in terms of the numbers of hidden layers and nodes in each layer. According to previous studies, one hidden layer is sufficient for approximating most continuous functions [33,34]. By contrast, NNs with more than one hidden layer are called deep NNs, and they have superior generalization ability to one-hidden-layer NNs [35,36]. With improvement of the hardware, deep learning models have become increasingly popular over the past few years. In this study, we constructed two types of NNs, namely a one-hidden-layer NN (NN1) and a five-hidden-layer NN (NN5), as our underlying architectures. NN1 represents the simplest form of NNs, whereas NN5 represents a deep learning model.

To determine the appropriate architecture of an NN model, some previous studies offered theoretical heuristics regarding the number of hidden units in an NN layer. However, the results ranged widely according to different studies regarding the optimal number of nodes in a hidden layer [37-40]. Accordingly, in this study, we selected the final number of hidden units according to the cross-validation results. After hyperparameter tuning, we determined that the final sizes of the NN1 and NN5 architectures were 268×2048×1 and 268×1024×1024×1024×1024×1024×1, respectively, for the mass spectrometry data set and 9×128×1 and 9×512×512×512×512×1, respectively, for the urinalysis data set.

Pruned NNs

Pruning is a method for eliminating redundant connections in NNs [41]. In this method, an NN is converted into a sparse model to reduce its size. Pruning methods can be unstructured or structured [42]. Unstructured pruning eliminates the individual parameters in an NN, whereas structured pruning eliminates the connections in large units such as hidden units in a fully connected layer or channels in a convolutional layer.

In this study, we applied global unstructured pruning to eliminate connections from the entire NN. The pruned NNs displayed in the figures are NN5s with a sparsity of 50%. However, we implemented pruning with sparsity values of 25%, 50%, and 75% for the NN1 and NN5 models. The detailed results regarding other pruned NNs are provided in Multimedia Appendix 2-5.

Quantized NNs

Quantization is a common method for model compression. In this method, the model size is reduced by computing and storing parameters with low bit widths [43]. Two main quantization methods exist in the Pytorch framework: dynamic and static quantization [22]. Dynamic quantization is the simplest quantization method. In dynamic quantization, the weights of the quantized layers in an NN are replaced with low-precision data, and the activations are quantized just before entering each quantized layer during inference. By contrast, in static quantization, the parameters for activation quantization are determined before the inference phase. Therefore, static quantization requires an additional calibration with a data set before inference.

In this study, the parameters of the original NN1 and NN5 models were tensors in the single-precision floating-point format; the quantized models had a quantized 8-bit signed integer data format. The quantized NNs displayed in the figures are NN5s. However, we implemented dynamic quantization for both the NN1 and NN5 models, and the detailed results regarding quantized NNs are provided in Multimedia Appendix 2-5.

Model Construction

In this study, we selected the aforementioned supervised ML algorithms according to their maturity and popularity. For every ML model, we tuned the hyperparameters in each algorithm through 5-fold cross-validation. The cutoff with the highest Youden index was selected as the final cutoff in each model [44].

Model Comparison on Deployment

Predictive Performance

We evaluated the predictive performance of all final models using independent testing data sets. We selected accuracy and AUROC as the predictive performance metrics. The 95% CIs of both accuracy and AUROC were calculated.

Time and Power Consumption

We derived the inference time and power data from Intel Power Gadget 3.5 [23]. This commercial product provides power data on the basis of Intel Running Average Power Limit (RAPL) interface estimation. RAPL is a driver that provides a set of performance counter data on time, power, and energy [45,46].

We implemented the ML models using command lines and logged the time and power data using PowerLog3.0.exe, a command line version of Intel Power Gadget that allows users to log the time and power data of a specific command line. In addition, because Intel Power Gadget only provides the energy data of the entire processor, all testing procedures were performed without background programs. The measurement for each algorithm was repeated 100 times.

Statistical Analysis

We initially employed the Shapiro-Wilk test to check for normality. If the assumption of normality did not hold, we subsequently adopted the Friedman test to compare the means of different groups. The pairwise Wilcoxon signed-rank test was used to identify which groups were different. *P* values were adjusted using the Bonferroni multiple testing correction method. All statistical tests were two-sided with an α error level of .05.

Results

Predictive Performance of ML Algorithms

Figure 2 and Figure 3 display the classification accuracy rates and AUROC values for the various ML models, respectively. Almost all models had high accuracy rates. All algorithms, except for the kNN algorithm, achieved an accuracy rate of at least 70% for the mass spectrometry data set. Moreover, all algorithms, except for the SVM algorithm, achieved an accuracy rate of at least 70% for the urinalysis data set. As displayed in Figure 3, the two tree-based methods, namely the RF and XGB algorithms, achieved the two highest AUROC values for both datasets (84.7% and 83.9% for the mass spectrometry data set, respectively; 91.1% and 91.4% for the urinalysis data set, respectively). In particular, the RF and XGB algorithms exhibited significantly higher AUROC values than those of most of the other algorithms (eg, kNN, SVM, pruned five-hidden layer NN [PNN], and NN5) for the urinalysis data set. The results regarding the algorithms' predictive performance are detailed in Multimedia Appendix 2 and Multimedia Appendix 3.

Figure 2. Classification accuracy rates of different algorithms implemented on the mass spectrometry and urinalysis data sets. The black bars indicate the 95% CIs of the classification accuracy. LR: logistic regression; kNN: k-nearest neighbor; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting; NN1: one-hidden-layer neural network; QNN: quantized five-hidden-layer neural network; PNN: pruned five-hidden-layer neural network; NN5: five-hidden-layer neural network.

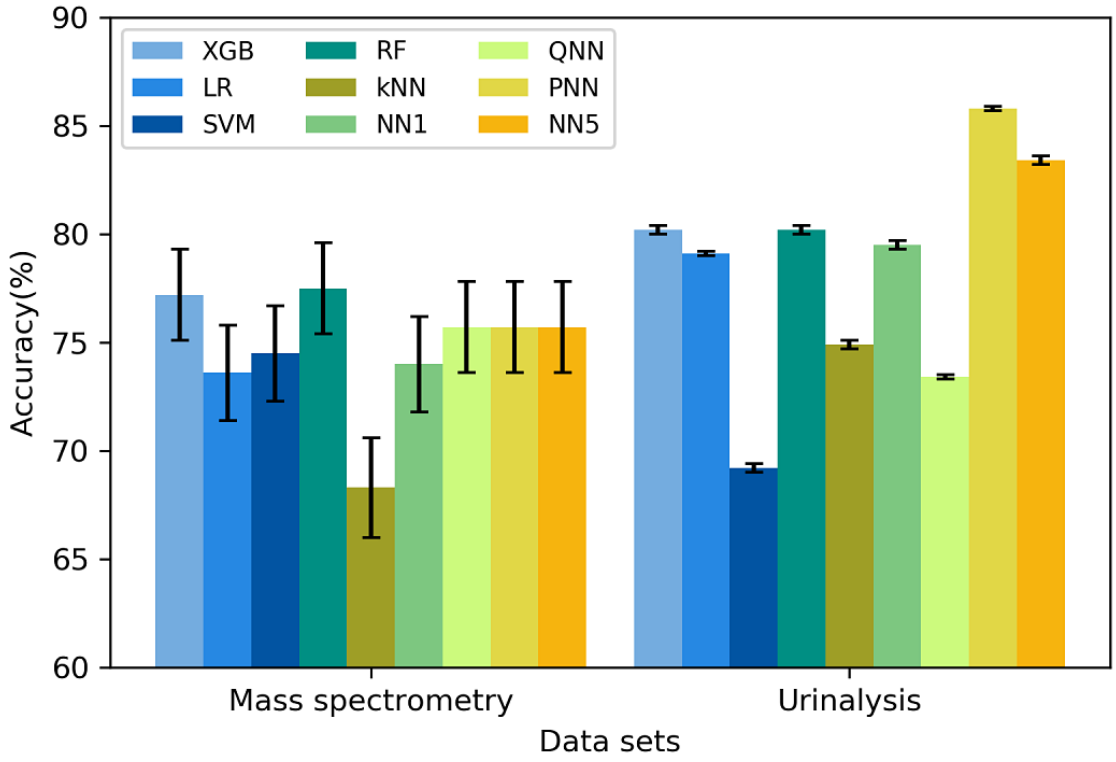
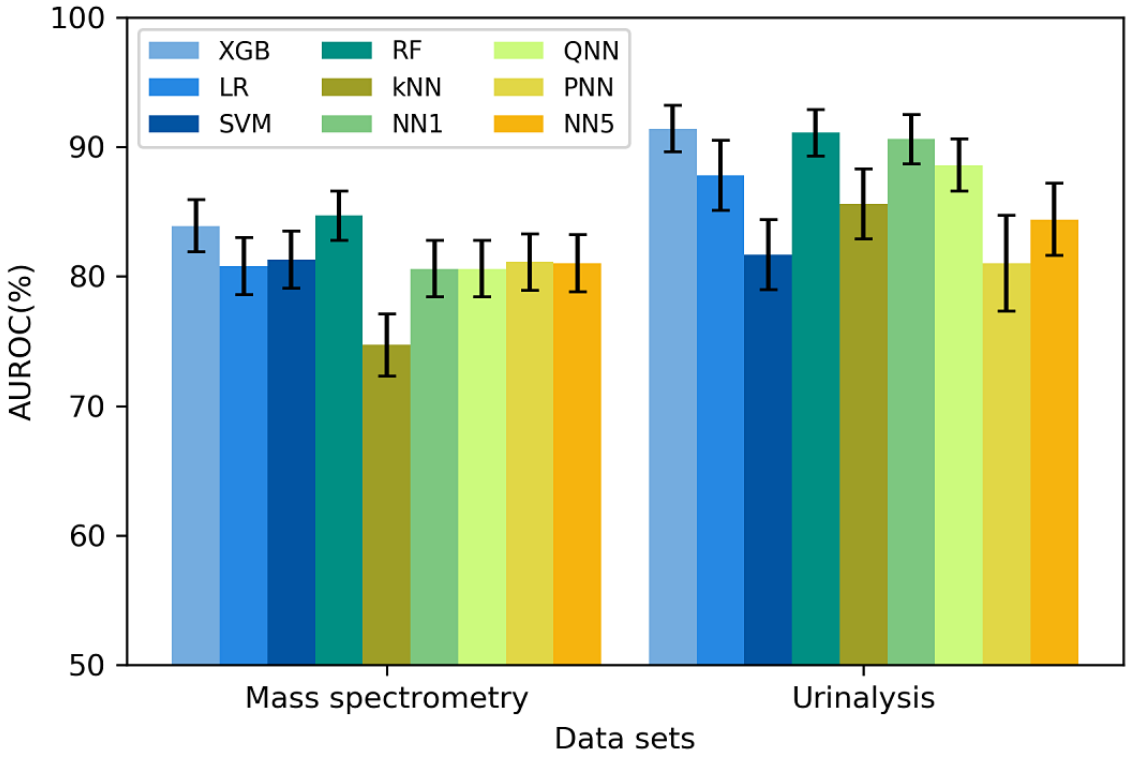


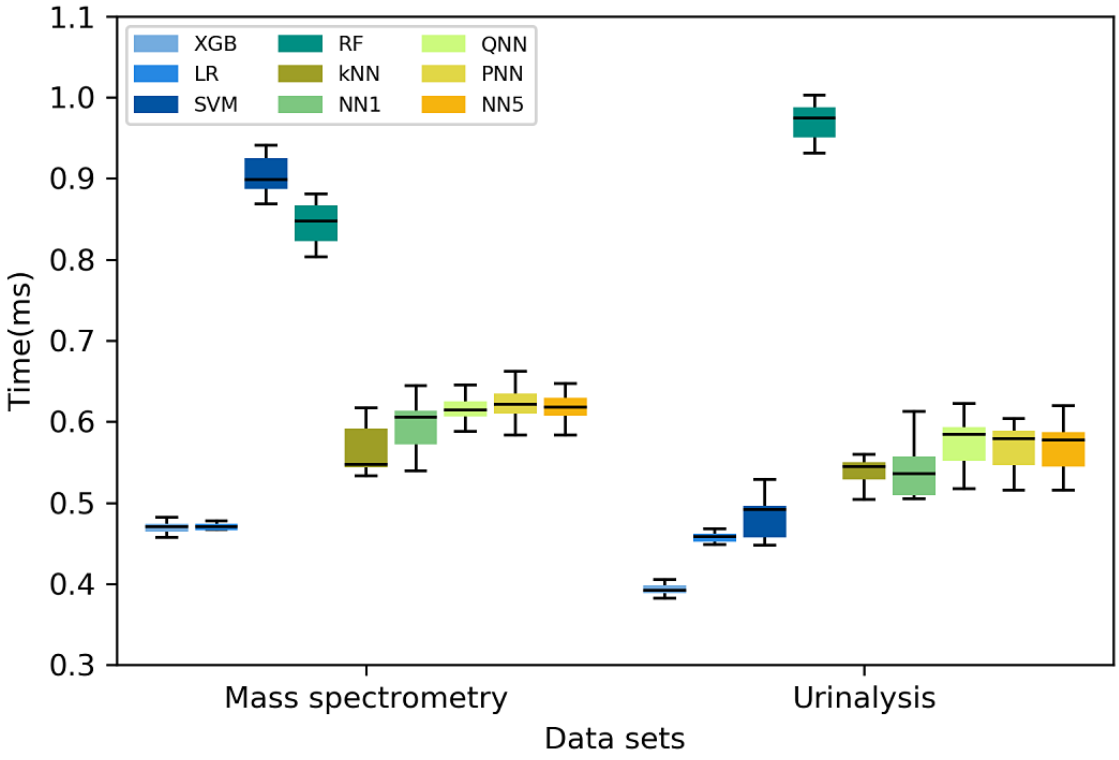
Figure 3. AUROC values of different algorithms implemented on the mass spectrometry and urinalysis data sets. The black bars indicate the 95% CIs of the AUROC. LR: logistic regression; kNN: k-nearest neighbor; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting; NN1: one-hidden-layer neural network; QNN: quantized five-hidden-layer neural network; PNN: pruned five-hidden-layer neural network; NN5: five-hidden-layer neural network; AUROC: area under the receiver operating characteristic curve.



Inference Times of ML Algorithms

Figure 4 presents a comparison of the inference times of the various ML algorithms. All algorithms completed the inference process within 1 millisecond. The XGB and LR algorithms had the shortest runtimes (0.47 milliseconds for both in the mass spectrometry data set; 0.39 and 0.47 milliseconds, respectively, for the urinalysis data set). The Wilcoxon signed-rank test results revealed that the run times of these two algorithms differed significantly (P<.001) from those of the other algorithms, except for NN1. The SVM and RF algorithms exhibited the highest time consumption for the mass spectrometry and urinalysis data sets, respectively. In particular, the RF algorithm exhibited a higher run time compared with that of all other algorithms, except for the SVM algorithm, for both data sets (P<.001). The results regarding the time consumption of the algorithms are detailed in Multimedia Appendix 4-5, and the corresponding P values derived from the Wilcoxon signed-rank test are presented in Multimedia Appendix 6-7.

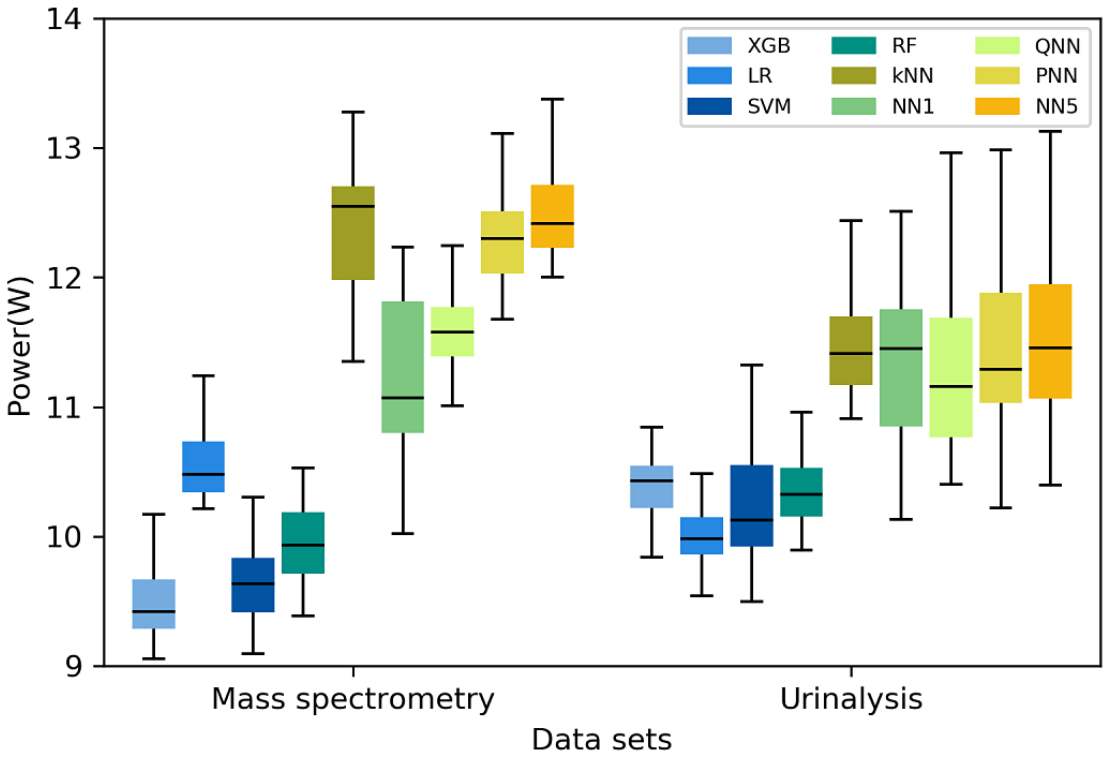
Figure 4. Time consumed in single prediction for the mass spectrometry and urinalysis data sets. LR: logistic regression; kNN: k-nearest neighbor; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting; NN1: one-hidden-layer neural network; QNN: quantized five-hidden-layer neural network; PNN: pruned five-hidden-layer neural network; NN5: five-hidden-layer neural network.



Power Consumption of ML Algorithms

Figure 5 presents a comparison of the power consumption levels of the ML algorithms. Algorithms of the same type consumed similar amounts of power. For example, both tree-based algorithms (RF and XGB) consumed limited power, whereas all NN-based models (NN1, quantized five-layer hidden NN [QNN], PNN, and NN5) consumed considerable power. The XGB algorithm exhibited the lowest power consumption for the mass spectrometry data set (9.42 Watts) and the LR algorithm exhibited the lowest power consumption for the urinalysis data set (9.98 Watts). According to the results of the Wilcoxon signed-rank tests (Multimedia Appendix 6-7), the LR and XGB algorithms exhibited lower power consumption levels than did the kNN algorithm and all NN-based algorithms for both datasets (P≤.001). The NN5, kNN, PNN, QNN, and NN1 algorithms exhibited higher power consumption levels compared with those of the other algorithms. Although pruning and quantization reduced the power consumption levels of the NN algorithms, the energy efficiency levels of the PNN and QNN algorithms did not surpass those of all the non-NN–based algorithms, except for the kNN algorithm. The results regarding power consumption are detailed in Multimedia Appendix 4-5, and the corresponding P values derived from the Wilcoxon signed-rank test are presented in Multimedia Appendix 6-7.

Figure 5. Power consumption levels of the different algorithms implemented on the mass spectrometry and urinalysis data sets. LR: logistic regression; kNN: k-nearest neighbor; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting; NN1: one-hidden-layer neural network; QNN: quantized five-hidden-layer neural network; PNN: pruned five-hidden-layer neural network; NN5: five-hidden-layer neural network.



Overall Comparison

Figure 6 and Figure 7 display scatter plots of the performance of the various algorithms in predicting *S. aureus* methicillin resistance and *T. vaginalis* infection. The horizontal and vertical axes in these figures represent the AUROC and average power consumption, respectively. The two dashed lines in the figures represent the average AUROC values and mean power consumption levels for the nine algorithms. Only the XGB and RF algorithms had higher than average AUROC and power consumption results for both data sets. Figure 6 and Figure 7 also illustrate the difference between the NN-based and non-NN–based algorithms. All NN-based algorithms are located in the lower half-plane in these figures and all the other algorithms, except for the kNN algorithm, are located in the upper half-plane in the figures. These results indicate that the NN-based algorithms had higher power consumption levels than those of the non-NN–based algorithms, even when model compression was executed through methods such as pruning or quantization.

Figure 6. Predictive performance (AUROC)–power consumption plot of the nine algorithms for the mass spectrometry data set. The two tree-based algorithms (RF and XGB) achieved a balanced predictive performance and power consumption. The horizontal and vertical dashed axes indicate the mean energy consumption and mean AUROC of the nine predictive models, respectively. Each algorithm is located in one of the four quadrants. The gray rectangle around each data point denotes the 95% CI of the AUROC and power consumption. LR: logistic regression; kNN: k-nearest neighbor; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting; NN1: one-hidden-layer neural network; QNN: quantized five-hidden-layer neural network; PNN: pruned five-hidden-layer neural network; NN5: five-hidden-layer neural network; AUROC: area under the receiver operating characteristic curve.

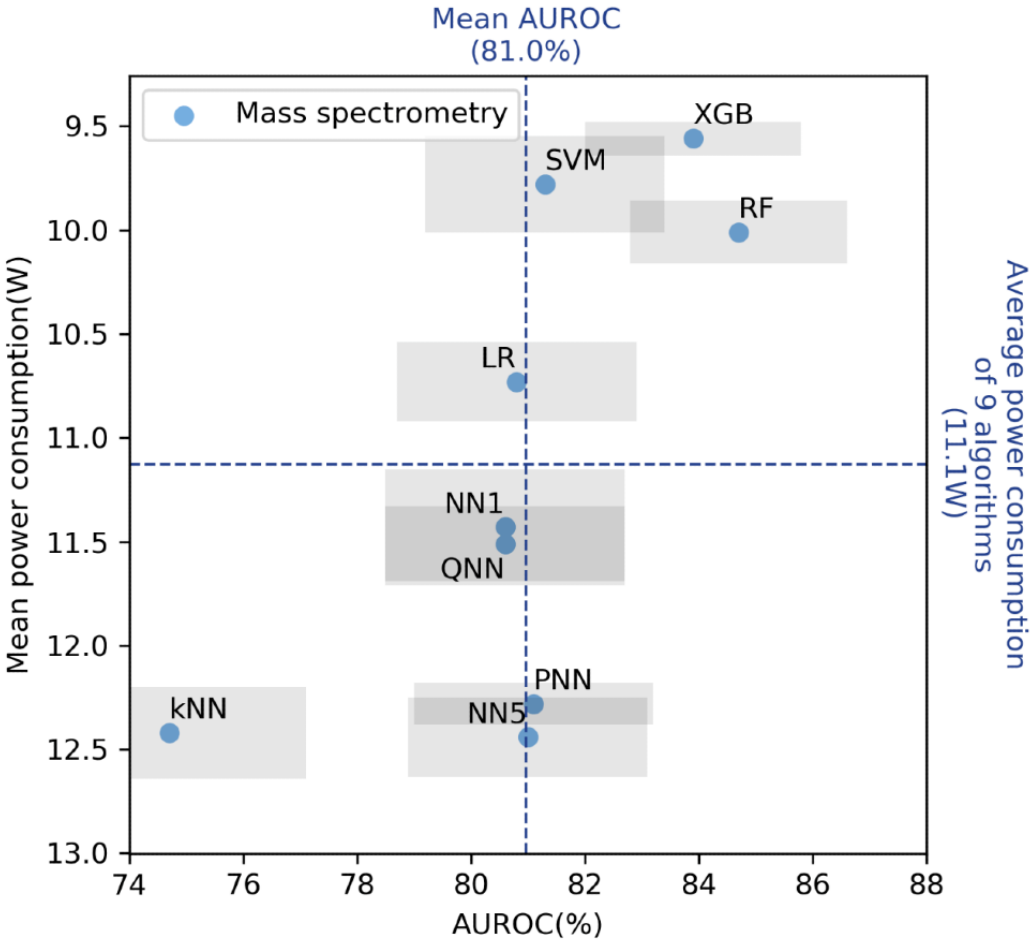
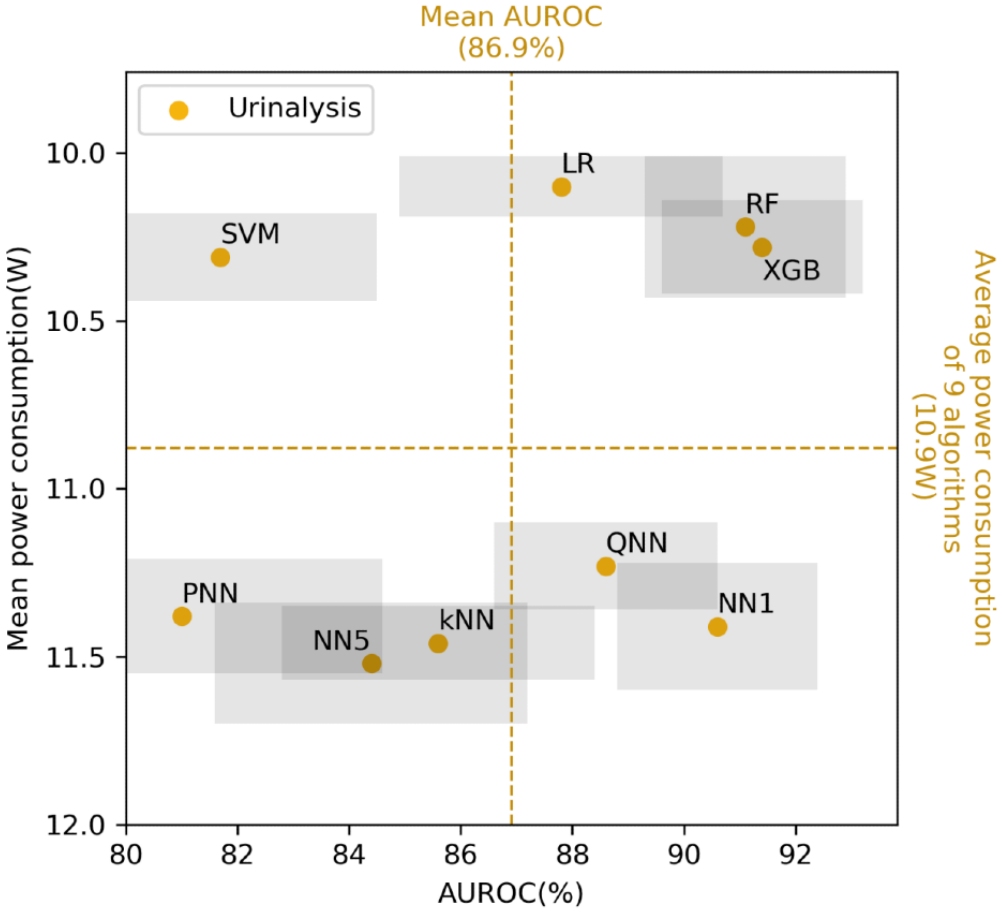


Figure 7. Predictive performance (AUROC)–power consumption plot of the nine algorithms for the urinalysis dataset. The two tree-based algorithms (ie, RF and XGB) achieved a balanced predictive performance and power consumption. The horizontal and vertical dashed axes indicate the mean energy consumption and mean AUROC of the nine predictive models, respectively. Each algorithm is located in one of the four quadrants. The gray rectangle around each data point denotes the 95% CI of the AUROC and power consumption. LR: logistic regression; kNN: k-nearest neighbor; SVM: support vector machine; RF: random forest; XGB: extreme gradient boosting; NN1: one-hidden-layer neural network; QNN: quantized five-hidden-layer neural network; PNN: pruned five-hidden-layer neural network; NN5: five-hidden-layer neural network; AUROC: area under the receiver operating characteristic curve.



Discussion

Principal Findings and Related Works

In this study, we compared the predictive performance, time consumption, and power consumption of nine algorithms using two clinical laboratory data sets. The XGB algorithm achieved a balanced performance with respect to the aforementioned metrics, indicating that the XGB algorithm is ideal for medical artificial intelligence applications with energy constraints.

In addition to this study, previous studies have performed comparative analyses of various ML algorithms in the medical domain [12,13,47]. However, only few studies have considered the inference efficiency in addition to the predictive performance. Zhang et al [13] compared the simplicity of seven algorithms by assessing their memory usage and training time for 12 public biomedical data sets. In another study, Deng et al [47] assessed the inference time of decision tree, SVM, RF, and NN algorithms. In this study, we executed our efficiency evaluation by directly exploring and comparing the power consumption levels of ML algorithms. Furthermore, all power consumption data were obtained according to real-time experimental results from performance counters.

Predictive Performance of ML Algorithms

The RF and XGB algorithms exhibited higher AUROC values than did the other algorithms for both data sets. This finding is similar to those of previous studies. In a study that considered 11 performance metrics, the RF algorithm and probability-calibrated boosted trees exhibited the best performance among 10 algorithms [48]. Other previous analyses also indicated that the RF and XGB algorithms consistently exhibit good performance for most biomedical data sets [13,14]. These algorithms have certain advantages; for example, they exhibit adequate scalability to large data sets and are more robust than other types of algorithms [17]. Medical data sets usually comprise features with strong signals; this is because only well-validated markers are routinely tested in clinical scenarios. Under this condition, tree-based methods would not be inferior to relatively complex models such as NN-based models. However, one should remember the “no free lunch theorem” [49], which suggests that no model exhibits superior performance universally. This statement is true because every algorithm is proposed on the basis of different underlying assumptions, which may fit only specific types of data. Therefore, different algorithms should be investigated when the

predictive performance of a certain model does not match the expectation.

Inference Time of ML Algorithms

In this study, the XGB and LR algorithms exhibited the shortest run times (both 0.47 milliseconds for the mass spectrometry data set; 0.39 and 0.47 milliseconds, respectively, for the urinalysis data set). The SVM and RF algorithms exhibited the highest time consumption levels for the mass spectrometry and urinalysis data sets, respectively. Notably, although the XGB and RF algorithms are ensemble algorithms based on decision trees, the XGB algorithm consumed less time than the RF algorithm. This finding is possibly due to differences in the depth and number of trees between these algorithms. For both data sets, the XGB model had shallower trees than did the RF model (for the mass spectrometry and urinalysis data sets, the maximum depths of the XGB decision trees were 6 and 10, respectively, and the average depths of the RF trees were 29 and 21, respectively). An explanation for this finding is that boosting reduces the bias of weak classifiers [17,50] and that bagging reduces the variance of complex classifiers [51]. Thus, the XGB algorithm may have shallower decision trees compared with those of the RF algorithm for the same prediction task. In addition to the depth difference, the number of trees may be another cause of the run time difference between the two tree-based algorithms (for the mass spectrometry and urinalysis data sets, the XGB algorithm contained 120 and 32 decision trees, respectively, and the RF algorithm contained more than 1000 decision trees). In an RF model, increasing the number of decision trees does not engender overfitting [18,30]. However, this characteristic may result in a final model with excessive decision trees after conventional grid-search cross-validation. By contrast, because an excessive number of decision trees results in overfitting in an XGB model, an XGB model with optimal predictive performance would have an appropriate number of trees. Furthermore, to identify the suitable tree numbers, the early stopping technique is frequently used during training of XGB models in practice [52]. In conclusion, the shorter run time of the XGB algorithm compared with the RF algorithm is possibly due to the different characteristics of these algorithms.

Power Consumption of ML Algorithms

The NN algorithms (NN1, QNN, PNN, and NN5) and the kNN algorithm exhibited the highest power consumption levels in this study, and the two tree-based algorithms (ie, RF and XGB) exhibited the lowest power consumption levels. Tree-based algorithms use the data structure of search trees for making inferences. The inference process mainly involves comparison operations at tree nodes and irregular memory access operations for subtree retrievals. In contrast to several other ML algorithms, tree-based algorithms typically do not use multiplication operations. The comparison and memory access operations in tree-based algorithms consume less energy than do multiplication operations [53,54]. The experiments in this study were run on a general-purpose CPU. Therefore, if necessary, the energy efficiency of tree-based algorithms can be increased using specialized hardware accelerations [55-57].

NNs have been regarded as the main tools for implementing ML in the last few years. The development of different NN architectures (eg, convolutional NNs and recurrent NNs) has contributed to considerable improvements in unstructured data analyses [35,55]. However, NNs have high power consumption. Thus, NNs should not always be considered as the preferred algorithm for implementing ML, unless they exhibit superior predictive performance compared with other algorithms. In this study, the adopted NNs consumed considerable power because of their high computational and communication demands. The computational demand of an NN refers to the large number of multiply-add operations in the forward propagation process, and the communicational demand of an NN refers to the energy cost of moving large quantities of data frequently between the processor and memory [7,58].

Several methods are available for reducing the power consumption of NNs. NNs have diverse architectures, and constructing an NN with a small architecture is an effective method for improving energy efficiency, as reflected by the difference in power consumption between the NN1 and NN5 models in this study (see Multimedia Appendix 3 and 5). In addition to constructing a small model, a given NN model can be compressed to reduce power consumption. In this study, we implemented and evaluated two common methods for NN compression, namely pruning [10,41] and quantization [59]. According to the obtained results, these model compression methods reduced the power consumption levels of the NNs. However, the NN-based algorithms did not exhibit higher energy efficiency levels compared with those of the non-NN–based algorithms, even after model compression. Furthermore, although energy optimization methods such as quantization are frequently used for NNs, these methods are not specific to NNs [60,61]. Thus, quantization can be feasibly applied to other ML algorithms if their power consumption must be decreased.

Overall Comparison

In summary, the XGB algorithm achieved balanced predictive performance and energy efficiency levels. Figure 6 and Figure 7 display the predictive performance–power consumption plots of the nine algorithms for the mass spectrometry and urinalysis data sets, respectively. In these figures, the two tree-based algorithms, namely the XGB and RF algorithms, are located in the right-upper quadrant, which indicates that they had higher than average predictive performance and lower than average power consumption. However, the XGB algorithm consumed less time than the RF algorithm ($P<.001$, according to the Wilcoxon signed-rank test; Figure 4 and Multimedia Appendix 6-7). Thus, the XGB algorithm achieved a higher energy efficiency level than the RF algorithm because the overall energy consumption for ML inference depends on not only power consumption but also on inference time.

Deep learning models are the main ML algorithms applied currently. These algorithms achieve state-of-the-art predictive performance for unstructured data sets (eg, data sets for computer vision and natural language processing) [55]. However, deep learning algorithms may be unnecessary for making predictions based on clinical laboratory data sets. In Figures 6 and 7, all of the NN-based algorithms are located in

the lower half-plane, signifying that the NN-based algorithms consumed more power than did most of the other algorithms. Pruning and quantization increased the efficiency levels of the NN-based algorithms; however, the increase was limited, and the energy efficiency levels of these algorithms did not surpass that of the XGB algorithm. Moreover, the NN-based algorithms did not exhibit higher AUROC values compared with those of the simple tree-based algorithms. The experimental results indicate that for data analysis in the clinical laboratory domain, simpler models such as the XGB model may be sufficient to achieve state-of-the-art predictive performance. Deep NNs are unsuitable for such data sets due to the high power consumption of these networks.

Limitations

This study has some limitations. First, because Intel Power Gadget 3.5 only provides the energy consumption of the entire processor [15], one should focus on the comparison of the investigated ML algorithms and not on the absolute power consumption obtained. Second, this study considered only two clinical laboratory data sets. Because energy consumption varies between data sets, a large-scale study based on a variety of medical data sets is essential for confirming the results of this study. Finally, the results were obtained using a general-purpose CPU; however, energy consumption may vary across different processors. Currently, ML is frequently implemented using hardware acceleration techniques. Although hardware devices such as discrete graph processing units or tensor processing units are not ubiquitous equipment in clinical settings, their energy efficiency levels are worth investigation. Energy efficiency is a major issue in embedded systems, and studies have been performed on the energy optimization of different algorithms [6,11,19]. Executing a fair comparison of energy efficiency under different hardware implementations is difficult. Hence, a well-designed comparative analysis of energy efficiency across different optimized methods is essential for obtaining general conclusions.

Conclusions

This study comprehensively compared various ML algorithms in terms of their predictive performance, time consumption, and power consumption when implemented on two clinical laboratory data sets. According to the results, the XGB algorithm attained balanced performance levels in terms of the aforementioned parameters for the two data sets. Thus, the XGB algorithm is ideal for application in real-world clinical settings.

Acknowledgments

This work was supported by Chang Gung Memorial Hospital (Linkou) (CMRPG3J1791, CMRPG3L0401, CMRPG3L0431, and CMRPG3L1011) and Ministry of Science and Technology, Taiwan (MOST 110-2636-E-008-008). This manuscript was edited by Wallace Academic Editing.

Authors' Contributions

HW conceptualized the study. JY, CHC, and TH wrote the manuscript, analyzed the data, plotted the figures, and created the tables. JY performed the experiments. TH, JL, CRC, TL, MW, YT, and HW reviewed and edited the manuscript for important intellectual content. YT and HW obtained funding and supervised the study. All authors discussed the results and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Time complexity of some common algorithms.
[DOCX File , 19 KB - jmir_v24i1e28036_app1.docx]

Multimedia Appendix 2
Classification performance of nonneural network–based machine learning algorithms implemented on the mass spectrometry and urinalysis data sets.
[DOCX File , 18 KB - jmir_v24i1e28036_app2.docx]

Multimedia Appendix 3
Classification performance of different neural networks (NNs) implemented on the mass spectrometry and urinalysis data sets.
[DOCX File , 16 KB - jmir_v24i1e28036_app3.docx]

Multimedia Appendix 4
Inferencing time and average power consumption levels of nonneural network–based algorithms implemented on the mass spectrometry and urinalysis data sets.
[DOCX File , 15 KB - jmir_v24i1e28036_app4.docx]

Multimedia Appendix 5
Inferencing time and average power consumption levels of different neural networks (NNs) implemented on the mass spectrometry and urinalysis data sets.
[DOCX File , 15 KB - jmir_v24i1e28036_app5.docx]

Multimedia Appendix 6
P values were derived from the pairwise Wilcoxon signed-rank test to identify which time and power consumption of any two algorithms were different on the mass spectrometry data set. The *P* values were adjusted by the Bonferroni multiple testing correction method. LR, logistic regression; kNN, k-nearest neighbors; SVM, support vector machine; RF, random forest; XGB, extreme gradient boosting; NN1, one-hidden-layer neural network; QNN, quantized five-hidden-layer neural network; PNN, pruned five-hidden-layer neural network; NN5, five-hidden-layer neural network.
[DOCX File , 15 KB - jmir_v24i1e28036_app6.docx]

Multimedia Appendix 7
P values were derived from the pairwise Wilcoxon signed-rank test to identify which time and power consumption of any two algorithms were different on the Urinalysis dataset. The adjusted *P* values was adjusted by the Bonferroni multiple testing correction method. LR, logistic regression; kNN, k-nearest neighbors; SVM, support vector machine; RF, random forest; XGB, extreme gradient boosting; NN1, one-hidden-layer neural network; QNN, quantized five-hidden-layer neural network; PNN, pruned five-hidden-layer neural network; NN5, five-hidden-layer neural network.
[DOCX File , 15 KB - jmir_v24i1e28036_app7.docx]

References

1. Wang H, Chen C, Shi S, Chung C, Wen Y, Wu M, et al. Improving multi-tumor biomarker health check-up tests with machine learning algorithms. Cancers (Basel) 2020 Jun 01;12(6):1442 [FREE Full text] [doi: 10.3390/cancers12061442] [Medline: 32492934]
2. Wang H, Hsieh C, Wen C, Wen Y, Chen C, Lu J. Cancers screening in an asymptomatic population by using multiple tumour markers. PLoS One 2016;11(6):e0158285 [FREE Full text] [doi: 10.1371/journal.pone.0158285] [Medline: 27355357]
3. Wang H, Chen C, Lee T, Horng J, Liu T, Tseng Y, et al. Rapid detection of heterogeneous vancomycin-intermediate based on matrix-assisted laser desorption ionization time-of-flight: using a machine learning approach and unbiased validation. Front Microbiol 2018;9:2393. [doi: 10.3389/fmicb.2018.02393] [Medline: 30364336]
4. Chatterjee A, Gerdes MW, Martinez SG. Identification of risk factors associated with obesity and overweight-a machine learning overview. Sensors (Basel) 2020 May 11;20(9):2734 [FREE Full text] [doi: 10.3390/s20092734] [Medline: 32403349]
5. Kohli M, Prevedello LM, Filice RW, Geis JR. Implementing machine learning in radiology practice and research. AJR Am J Roentgenol 2017 Apr;208(4):754-760. [doi: 10.2214/AJR.16.17224] [Medline: 28125274]
6. Kang M, Gonugondla S, Shanbhag N. A 19.4 nJ/decision 364K decisions/s in-memory random forest classifier in 6T SRAM array. 2017 Presented at: ESSCIRC 2017-43rd IEEE European Solid State Circuits Conference; September 11-14, 2017; Leuven, Belgium. [doi: 10.1109/esscirc.2017.8094576]
7. Rouhani B, Mirhoseini A, Koushanfar F. DeLight: adding energy dimension to deep neural networks. 2016 Presented at: 2016 International Symposium on Low Power Electronics and Design; August 8-10, 2016; San Francisco, CA. [doi: 10.1145/2934583.2934599]
8. Shoaib M, Jha N, Verma N. A low-energy computation platform for data-driven biomedical monitoring algorithms. 2011 Presented at: The 48th Annual Design Automation Conference 2011; June 5-10, 2011; San Diego, CA. [doi: 10.1145/2024724.2024861]
9. Gauen K, Rangan R, Mohan A, Lu Y, Liu W, Berg A. Low-power image recognition challenge. 2017 Presented at: 2017 22nd Asia and South Pacific Design Automation Conference; January 16-19, 2017; Chiba, Japan. [doi: 10.1109/aspdac.2017.7858303]
10. Yang T, Chen Y, Sze V. Designing energy-efficient convolutional neural networks using energy-aware pruning. 2017 Presented at: 2017 IEEE Conference on Computer Vision and Pattern Recognition; July 21-26, 2017; Honolulu, HI. [doi: 10.1109/cvpr.2017.643]
11. Ayinala M, Parhi K. Low-energy architectures for support vector machine computation. 2013 Presented at: 2013 Asilomar Conference on Signals, Systems and Computers; November 3-6, 2013; Pacific Grove, CA. [doi: 10.1109/acssc.2013.6810693]
12. Uddin S, Khan A, Hossain ME, Moni MA. Comparing different supervised machine learning algorithms for disease prediction. BMC Med Inform Decis Mak 2019 Dec 21;19(1):281 [FREE Full text] [doi: 10.1186/s12911-019-1004-8] [Medline: 31864346]
13. Zhang Y, Xin Y, Li Q, Ma J, Li S, Lv X, et al. Empirical study of seven data mining algorithms on different characteristics of datasets for biomedical classification applications. Biomed Eng Online 2017 Nov 02;16(1):125 [FREE Full text] [doi: 10.1186/s12938-017-0416-x] [Medline: 29096638]

14. Harper PR. A review and comparison of classification algorithms for medical decision making. Health Policy 2005 Mar;71(3):315-331. [doi: 10.1016/j.healthpol.2004.05.002] [Medline: 15694499]
15. García-Martín E, Rodrigues CF, Riley G, Grahn H. Estimation of energy consumption in machine learning. J Parallel Distrib Comput 2019 Dec;134:75-88. [doi: 10.1016/j.jpdc.2019.07.007]
16. Kibriya A, Frank E. An empirical comparison of exact nearest neighbour algorithms. 2007 Presented at: European Conference on Principles of Data Mining and Knowledge Discovery; September 17-21, 2007; Warsaw, Poland. [doi: 10.1007/978-3-540-74976-9_16]
17. Murphy K. Machine learning: a probabilistic perspective. Cambridge, MA: MIT Press; 2012.
18. Louppe G. Understanding random forests: From theory to practice. arXiv. 2014. URL: https://arxiv.org/abs/1407.7502 [accessed 2021-03-19]
19. Zhang B, Davoodi A, Hu YH. Exploring energy and accuracy tradeoff in structure simplification of trained deep neural networks. IEEE J Emerg Sel Topics Circuits Syst 2018 Dec;8(4):836-848. [doi: 10.1109/jetcas.2018.2833383]
20. Abraham A, Pedregosa F, Eickenberg M, Gervais P, Mueller A, Kossaifi J, et al. Machine learning for neuroimaging with scikit-learn. Front Neuroinform 2014;8:14. [doi: 10.3389/fninf.2014.00014] [Medline: 24600388]
21. Chen T, Guestrin C. Xgboost: A scalable tree boosting system. 2016 Presented at: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; August 13-17, 2016; San Francisco, CA. [doi: 10.1145/2939672.2939785]
22. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G. Pytorch: An imperative style, high-performance deep learning library. arXiv. 2019. URL: https://arxiv.org/abs/1912.01703 [accessed 2021-07-31]
23. Intel® 64 and ia-32 architectures software developer's manual, Volume 3B: System Programming Guide, Part 2. Intel Corporation. 2011. URL: https://www.intel.com.tw/ [accessed 2021-03-19]
24. Wang H, Chung C, Wang Z, Li S, Chu B, Horng J, et al. A large-scale investigation and identification of methicillin-resistant Staphylococcus aureus based on peaks binning of matrix-assisted laser desorption ionization-time of flight MS spectra. Brief Bioinform 2021 May 20;22(3):bbaa138 [FREE Full text] [doi: 10.1093/bib/bbaa138] [Medline: 32672791]
25. Wang Z, Wang H, Chung C, Horng J, Lu J, Lee T. Large-scale mass spectrometry data combined with demographics analysis rapidly predicts methicillin resistance in Staphylococcus aureus. Brief Bioinform 2021 Jul 20;22(4):bbaa293. [doi: 10.1093/bib/bbaa293] [Medline: 33197936]
26. Wang H, Hung C, Chen C, Lee T, Huang K, Ning H, et al. Increase Trichomonas vaginalis detection based on urine routine analysis through a machine learning approach. Sci Rep 2019 Aug 19;9(1):11074. [doi: 10.1038/s41598-019-47361-8] [Medline: 31423009]
27. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002 Jun 01;16:321-357 [FREE Full text] [doi: 10.1613/jair.953]
28. Manning C, Schütze H, Raghavan P. Introduction to information retrieval. Cambridge, UK: Cambridge University Press; 2008.
29. Rashidi HH, Sen S, Palmieri TL, Blackmon T, Wajda J, Tran NK. Early recognition of burn- and trauma-related acute kidney injury: a pilot comparison of machine learning techniques. Sci Rep 2020 Jan 14;10(1):205. [doi: 10.1038/s41598-019-57083-6] [Medline: 31937795]
30. Flaxman AD, Vahdatpour A, Green S, James SL, Murray CJ, Population Health Metrics Research Consortium (PHMRC). Random forests for verbal autopsy analysis: multisite validation study using clinical diagnostic gold standards. Popul Health Metr 2011 Aug 04;9:29 [FREE Full text] [doi: 10.1186/1478-7954-9-29] [Medline: 21816105]
31. Couronné R, Probst P, Boulesteix A. Random forest versus logistic regression: a large-scale benchmark experiment. BMC Bioinformatics 2018 Jul 17;19(1):270 [FREE Full text] [doi: 10.1186/s12859-018-2264-5] [Medline: 30016950]
32. Schapire RE, Freund Y. Boosting: Foundations and Algorithms. Cambridge, MA: MIT Press; Jan 04, 2013:164-166.
33. Hecht-Nielsen R. On the algebraic structure of feedforward network weight spaces. In: Eckmiller R, editor. Advanced Neural Computers. North Holland: Elsevier; 1990:129-135.
34. Basheer IA, Hajmeer M. Artificial neural networks: fundamentals, computing, design, and application. J Microbiol Methods 2000 Dec 01;43(1):3-31. [doi: 10.1016/s0167-7012(00)00201-3] [Medline: 11084225]
35. Kriegeskorte N, Golan T. Neural network models and deep learning. Curr Biol 2019 Apr 01;29(7):R231-R236 [FREE Full text] [doi: 10.1016/j.cub.2019.02.034] [Medline: 30939301]
36. Schmidhuber J. Deep learning in neural networks: an overview. Neural Netw 2015 Jan;61:85-117. [doi: 10.1016/j.neunet.2014.09.003] [Medline: 25462637]
37. Widrow B, Lehr M. 30 years of adaptive neural networks: perceptron, Madaline, and backpropagation. Proc IEEE 1990 Sep;78(9):1415-1442. [doi: 10.1109/5.58323]
38. Masters T. Signal And Image Processing With Neural Networks: A C++ Sourcebook. Hoboken, NJ: John Wiley & Sons, Inc; 1994.
39. Lachtermacher G, Fuller JD. Back propagation in time-series forecasting. J Forecast 1995 Jul;14(4):381-393. [doi: 10.1002/for.3980140405]
40. Jadid MN, Fairbairn DR. Neural-network applications in predicting moment-curvature parameters from experimental data. Eng Appl Artif Intell 1996 Jun;9(3):309-319. [doi: 10.1016/0952-1976(96)00021-8]

41. Han S, Pool J, Tran J, Dally WJ. Learning both weights and connections for efficient neural networks. arXiv. 2015 Oct 30. URL: https://arxiv.org/abs/1506.02626 [accessed 2021-11-08]
42. Liu Z, Sun M, Zhou T, Huang G, Darrell T. Rethinking the value of network pruning. arXiv. 2018. URL: https://arxiv.org/abs/1810.05270 [accessed 2021-07-31]
43. Hubara I, Courbariaux M, Soudry D, El-Yaniv R, Bengio Y. Quantized neural networks: Training neural networks with low precision weights and activations. J Machine Learn Res 2017;18(1):6869-6898.
44. Krzanowski W, Hand D. ROC curves for continuous data. Boca Raton, FL: CRC Press; 2009.
45. David H, Gorbatov E, Hanebutte U, Khanna R, Le C. RAPL: Memory power estimation and capping. 2010 Presented at: 2010 ACM/IEEE International Symposium on Low-Power Electronics and Design; August 18-20, 2010; Austin, TX.
46. Czarnul P, Proficz J, Krzywaniak A. Energy-aware high-performance computing: survey of state-of-the-art tools, techniques, and environments. Sci Program 2019 Apr 24;2019:8348791 [FREE Full text] [doi: 10.1155/2019/8348791]
47. Deng F, Huang J, Yuan X, Cheng C, Zhang L. Performance and efficiency of machine learning algorithms for analyzing rectangular biomedical data. Lab Invest 2021 Apr 11;101(4):430-441. [doi: 10.1038/s41374-020-00525-x] [Medline: 33574440]
48. Caruana R, Niculescu-Mizil A. An empirical comparison of supervised learning algorithms. 2006 Presented at: 23rd International Conference on Machine Learning; June 25-29, 2006; Pittsburgh, PA. [doi: 10.1145/1143844.1143865]
49. Wolpert DH. The lack of a priori distinctions between learning algorithms. Neural Comput 1996 Oct;8(7):1341-1390. [doi: 10.1162/neco.1996.8.7.1341]
50. Lee H, Yoon H, Nam K, Cho YJ, Kim TK, Kim WH, et al. Derivation and validation of machine learning approaches to predict acute kidney injury after cardiac surgery. J Clin Med 2018 Oct 03;7(10):322 [FREE Full text] [doi: 10.3390/jcm7100322] [Medline: 30282956]
51. Breiman L. Bagging predictors. Mach Learn 1996 Aug;24(2):123-140. [doi: 10.1007/bf00058655]
52. Zhang T, Yu B. Boosting with early stopping: convergence and consistency. Ann Statist 2005 Aug 1;33(4):1538-1579. [doi: 10.1214/009053605000000255]
53. McKeown M, Lavrov A, Shahrad M, Jackson P, Fu Y, Balkind J. Power and energy characterization of an open source 25-Core Manycore processor. 2018 Presented at: 2018 IEEE International Symposium on High Performance Computer Architecture; February 24-28, 2018; Vienna, Austria. [doi: 10.1109/hpca.2018.00070]
54. Vasilakis E. An instruction level energy characterization of ARM processors. Technical Report FORTH-ICS/TR-450, March 2015. GreenVM. 2015 Mar. URL: https://projects.ics.forth.gr/carv/greenvm/files/tr450.pdf [accessed 2021-07-31]
55. Shoaran M, Haghi BA, Taghavi M, Farivar M, Emami-Neyestanak A. Energy-efficient classification for resource-constrained biomedical applications. IEEE J Emerg Sel Topics Circuits Syst 2018 Dec;8(4):693-707. [doi: 10.1109/jetcas.2018.2844733]
56. Takhirov Z, Wang J, Louis M, Saligrama V, Joshi A. Field of groves: an energy-efficient random forest. arXiv. 2017. URL: https://arxiv.org/abs/1704.02978 [accessed 2021-07-31]
57. Van Essen B, Macaraeg C, Gokhale M, Prenger R. Accelerating a random forest classifier: Multi-Core, GP-GPU, or FPGA? 2012 Presented at: 2012 IEEE 20th International Symposium on Field-Programmable Custom Computing Machines; April 29-May 1, 2012; Toronto, ON. [doi: 10.1109/fccm.2012.47]
58. Chen Y, Emer J, Sze V. Eyeriss. SIGARCH Comput Archit News 2016 Oct 12;44(3):367-379. [doi: 10.1145/3007787.3001177]
59. Gong Y, Liu L, Yang M, Bourdev L. Compressing deep convolutional networks using vector quantization. arXiv. 2014. URL: https://arxiv.org/abs/1412.6115 [accessed 2021-07-31]
60. Zhu B, Shoaran M. Hardware-efficient seizure detection. 2019 Presented at: 53rd Asilomar Conference on Signals, Systems, and Computers; November 3-6, 2019; Pacific Grove, CA. [doi: 10.1109/ieeeconf44664.2019.9049047]
61. Jégou H, Douze M, Schmid C. Product quantization for nearest neighbor search. IEEE Trans Pattern Anal Mach Intell 2011 Jan;33(1):117-128. [doi: 10.1109/tpami.2010.57]

Abbreviations

AUROC: area under the receiver operating characteristic curve
CPU: central processing unit
kNN: k-nearest neighbor
LR: logistic regression
ML: machine learning
NN: neural network
NN1: one-hidden-layer neural network
NN5: five-hidden-layer neural network
PNN: pruned five-hidden-layer neural network
QNN: quantized five-hidden-layer neural network
RAPL: Running Average Power Limit
RF: random forest

SVM: support vector machine
XGB: extreme gradient boosting

Edited by R Kukafka; submitted 19.03.21; peer-reviewed by A Chatterjee, J Yang; comments to author 29.04.21; revised version received 31.07.21; accepted 04.10.21; published 25.01.22.

Please cite as:
Yu JR, Chen CH, Huang TW, Lu JJ, Chung CR, Lin TW, Wu MH, Tseng YJ, Wang HY
Energy Efficiency of Inference Algorithms for Clinical Laboratory Data Sets: Green Artificial Intelligence Study
J Med Internet Res 2022;24(1):e28036
URL: https://www.jmir.org/2022/1/e28036
doi:10.2196/28036
PMID:35076405

Original Paper

Identifying Information Gaps in Electronic Health Records by Using Natural Language Processing: Gynecologic Surgery History Identification

Sungrim Moon[1], PhD; Luke A Carlson[1], BA; Ethan D Moser[2], BS; Bhavani Singh Agnikula Kshatriya[1], MS; Carin Y Smith[3], BS; Walter A Rocca[2,4,5], MD, MPH; Liliana Gazzuola Rocca[2], MD; Suzette J Bielinski[2], PhD; Hongfang Liu[1], PhD; Nicholas B Larson[3], PhD

[1]Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, United States

[2]Division of Epidemiology, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, United States

[3]Division of Clinical Trials and Biostatistics, Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, United States

[4]Department of Neurology, Mayo Clinic, Rochester, MN, United States

[5]Women's Health Research Center, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Nicholas B Larson, PhD
Division of Clinical Trials and Biostatistics
Department of Quantitative Health Sciences
Mayo Clinic
200 First Street SW
Rochester, MN, 55905
United States
Phone: 1 507 293 1700
Email: Larson.Nicholas@mayo.edu

Abstract

Background: Electronic health records (EHRs) are a rich source of longitudinal patient data. However, missing information due to clinical care that predated the implementation of EHR system(s) or care that occurred at different medical institutions impedes complete ascertainment of a patient's medical history.

Objective: This study aimed to investigate information discrepancies and to quantify information gaps by comparing the gynecological surgical history extracted from an EHR of a single institution by using natural language processing (NLP) techniques with the manually curated surgical history information through chart review of records from multiple independent regional health care institutions.

Methods: To facilitate high-throughput evaluation, we developed a rule-based NLP algorithm to detect gynecological surgery history from the unstructured narrative of the Mayo Clinic EHR. These results were compared to a gold standard cohort of 3870 women with gynecological surgery status adjudicated using the Rochester Epidemiology Project medical records–linkage system. We quantified and characterized the information gaps observed that led to misclassification of the surgical status.

Results: The NLP algorithm achieved precision of 0.85, recall of 0.82, and F1-score of 0.83 in the test set (n=265) relative to outcomes abstracted from the Mayo EHR. This performance attenuated when directly compared to the gold standard (precision 0.79, recall 0.76, and F1-score 0.76), with the majority of misclassifications being false negatives in nature. We then applied the algorithm to the remaining patients (n=3340) and identified 2 types of information gaps through error analysis. First, 6% (199/3340) of women in this study had no recorded surgery information or partial information in the EHR. Second, 4.3% (144/3340) of women had inconsistent or inaccurate information within the clinical narrative owing to misinterpreted information, erroneous "copy and paste," or incorrect information provided by patients. Additionally, the NLP algorithm misclassified the surgery status of 3.6% (121/3340) of women.

Conclusions: Although NLP techniques were able to adequately recreate the gynecologic surgical status from the clinical narrative, missing or inaccurately reported and recorded information resulted in much of the misclassification observed. Therefore, alternative approaches to collect or curate surgical history are needed.

(J Med Internet Res 2022;24(1):e29015) doi:10.2196/29015

KEYWORDS

information gap; health information interoperability; natural language processing; electronic health records; gynecologic surgery; surgery; medical informatics; digital health; eHealth; gynecology

Introduction

Electronic health records (EHRs) are a rich source of longitudinal patient information that can efficiently and cost-effectively be used for clinical care as well as for research. However, missing information due to clinical care that predated the implementation of EHR system(s) or that occurred at different medical institutions may result in an incomplete medical history. For example, gynecologic surgery history is essential for assessing women's health, given the increased risk of aging-related outcomes among women undergoing these surgeries [1-6]. However, assessment of surgical status is complicated by the significant time interval (ie, decades) between these procedures and the subsequent aging-related events because these procedures occurred at different medical institutions. In addition, collecting comprehensive gynecological surgery history is challenging because various surgical combinations are performed: hysterectomy with or without oophorectomy, unilateral or bilateral oophorectomy, and unilateral oophorectomy followed by the removal of the remaining ovary at a later date.

The approaches to mitigate these types of information gaps in a patient's medical history are (1) patient-provided information either by questionnaires or data collection during a clinical visit or (2) chart review. However, time constraints on providers can delay or prevent accurate assessment of medical history [7,8]. Further, patient-provided information can be limited or be inaccurate due to recall errors or lack of health literacy [9,10]. Manual chart abstraction of past medical records can overcome these issues but is often labor-intensive and time-consuming.

Natural language processing (NLP) techniques may be used to automatically extract relevant clinical information in a high throughput fashion. However, the medical history information of a patient is often a mix of paper records and EHRs distributed over multiple systems within or across multiple health care institutions [11]. This can be due to the evolution of clinical documentation at a single health care institution or the involvement of multiple health care institutions over the lifespan of patients. In some instances when upgrading EHR systems, past records are not loaded and some data elements may be completely dropped owing to differences in the underlying data models between the 2 systems [12,13]. In addition, patients can move in and out of health care institutions over time owing to personal preference, insurance coverage, or the referral process [14].

In this study, we had a unique opportunity to quantify information gaps by comparing the historical gynecologic surgery information obtained from EHR data of a single institution by using NLP techniques with the surgical history information that was manually curated through chart review of records from multiple independent regional health care institutions.

Methods

Gold Standard Cohort

The Mayo Clinic Cohort Study for Oophorectomy and Aging-2 (MOA-2) consisted of 570 women who underwent unilateral oophorectomy and 1653 women who underwent bilateral oophorectomy in Olmsted County, Minnesota between 1988 and 2007 before the age of 50 years [5,15,16]. Bilateral oophorectomy was defined as the removal of both ovaries in the same surgery or as the removal of the remaining ovary if 2 separate unilateral oophorectomies were performed. Women were excluded if they had undergone natural menopause before the oophorectomy. Women were also excluded if the oophorectomy was performed as a treatment for ovarian cancer, for estrogen-sensitive cancer, or if they carried a high-risk genetic variant. Each woman was matched by age (+/- 1 year) to a population-based referent woman who had not undergone any oophorectomy (570 unilateral referent women) or bilateral oophorectomy (1653 bilateral referent women) as of the date of surgery (index date) [5,15,16].

All women were identified using the Rochester Epidemiology Project (REP) medical records–linkage system [17-20]. Each health care provider in Olmsted County, Minnesota, uses a unit (or dossier) medical record system whereby all data collected on an individual are assembled in one place. Through the REP, these health care providers have agreed to share their patient records for research studies approved by the Institutional Review Boards of Mayo Clinic and Olmsted Medical Center [17]. In 2017, the REP contained approximately 2.3 million patient records from 54 different health care providers that matched to more than 591,000 individuals who had been residents of the Olmsted County at some point between 1966 and 2017. The REP captures virtually the entire population of Olmsted County as compared to the US Census (>99.9% of the 1970-2010 census counts) [18].

In MOA-2, available paper medical records and EHR data for each of the women were manually abstracted to confirm gynecological surgeries from all available REP sources before the index date and up to the last follow-up date. Thus, MOA-2 represents a gold standard data set with complete capture of surgical histories from all REP sources. Gynecological surgery status was divided into the following 6 mutually exclusive categories: bilateral oophorectomy only, hysterectomy and bilateral oophorectomy, unilateral oophorectomy only, hysterectomy and unilateral oophorectomy, hysterectomy only, and no surgery. Since each woman may have undergone multiple gynecological surgeries throughout her life (eg, an initial hysterectomy followed by a bilateral oophorectomy at a later date), a single status was assigned as of the latest individual

follow-up date for each woman. Follow-up dates ranged from January 1997 through August 2019.

MOA-2 included 4446 women, of whom 173 were represented in the cohort twice, leaving 4273 unique women. For this study, we excluded women who died prior to the start of the Mayo Clinic EHR in 1997 (n=13), women without a Mayo Clinic medical record number (n=28), women who did not provide research authorization for medical records review (n=102), or women with no information available in the Mayo Clinic EHR (n=260). The final cohort consisted of 3870 unique women.

Single-Institution Surgery Status Abstraction

Using labels from the gold standard, we randomly selected 100 women from each surgical status category for train and test sets (Table 1). However, owing to the rarity of "bilateral oophorectomy without hysterectomy," only 30 women were included. The surgical status was then reviewed for women included in the train and test sets (n=530) by one of the 2 trained annotators (EDM and Ellen E Koepsell) using only data available within the Mayo EHR. The annotators were blinded to both the external gold standard status and abstractions of the other annotator. A stratified random sample by surgery type of 10.2% (54/530) of the women was additionally used to assess interannotator reliability, which was evaluated by percentage agreement and Cohen kappa.

Table 1. Gynecological surgical status of the patients in this study (N=3870).

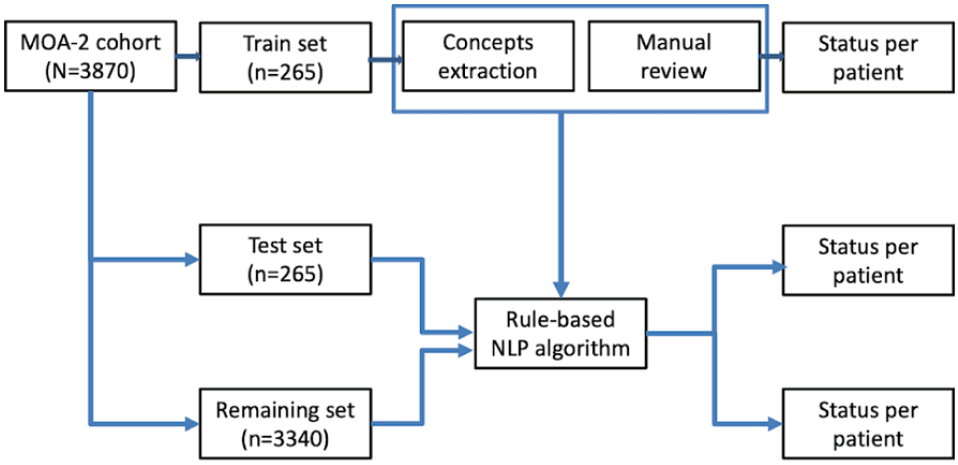
Surgery status in MOA-2[a]	Train set (n=265)	Test set (n=265)	Remaining set (n=3340)
No surgery (n=1473)	50	50	1373
Bilateral oophorectomy only (n=35)	15	15	5
Hysterectomy and bilateral oophorectomy (n=1685)	50	50	1585
Unilateral oophorectomy only (n=214)	50	50	114
Hysterectomy and unilateral oophorectomy (n=247)	50	50	147
Hysterectomy only (n=216)	50	50	116

[a]MOA-2: Mayo Clinic Cohort Study of Oophorectomy and Aging-2.

NLP

To facilitate high-throughput surgical status extraction from the Mayo Clinic EHR, the train set was used to develop a rule-based NLP algorithm, and the test set was used to evaluate the NLP algorithm performance (Figure 1). The NLP algorithm was built using MedTagger applied to text from clinical notes, as well as pathology, radiology, and surgical operative reports in the Mayo Clinic EHR. MedTagger is a pipeline tool capable of extracting clinical events from the unstructured text given a clinical dictionary and ruleset [21]. MedTagger was designated as an NLP platform by Mayo Clinic for clinical NLP research. To develop the NLP algorithm to determine the status of gynecological surgery for each woman, MedTagger was adapted to extract surgery concepts within the clinical sections relevant to medical history and current clinical care (Multimedia Appendix 1). In detail, we utilized the series of the pipeline of MedTagger, such as sentence detection, tokenization, concept identification, and assertion. We aggregated the extracted concepts based on rules (Multimedia Appendix 2) at the patient level to determine the status of the patient's surgery. For example, a sentence in clinical notes, "A total abdominal hysterectomy with bilateral salpingo-oophorectomy was performed the usual fashion," triggers 2 concepts, "hysterectomy" and "bilateral salpingo-oophorectomy," through the pipeline of MedTagger. The NLP algorithm determines the patient's surgery status as "Hysterectomy and bilateral oophorectomy." Only concepts relevant to the women (ie, not family history) with positive and assertive contextual information were considered valid. If the sentence included a valid oophorectomy concept and contained the word "left," this was categorized as "left side oophorectomy," whereas those having the word "right" were categorized as "right side oophorectomy." During the process of aggregating the extracted concepts on the patient level, if none of the concepts contain the laterality of a unilateral oophorectomy surgery, it was considered "left side oophorectomy" as default and classified as unilateral oophorectomy. The final surgical status for each woman was determined by applying rules to all valid concepts relevant to the woman (Multimedia Appendix 2). Because temporal information is also critical, we explored the extraction of the surgery date information for 3 types of surgeries, that is, unilateral oophorectomy, bilateral oophorectomy, and hysterectomy in the train and test sets (n=530). We extracted all date information based on 3 patterns, that is, DD/MM/YYYY, DD/MM/YY, or YYYY from sentences containing the surgery information.

Figure 1. An overview of this study to classify the surgical histories of patients. MOA-2: Mayo Clinic Cohort Study of Oophorectomy and Aging-2; NLP: natural language processing.



Performance Evaluation

To evaluate the performance of the NLP algorithm, we calculated precision (ie, positive predictive value), recall (ie, sensitivity), F1-score, and accuracy. Precision represented the proportion of women that the NLP algorithm determined as having surgery who truly had the surgery. Recall indicated the proportion of women who truly had surgery and were determined by the NLP algorithm as having had surgery. F1-score was the harmonized measurement between precision and recall. Accuracy was the proportion of correctly classified surgery statuses by the NLP algorithm. All performance measures were calculated both with respect to surgical status ascertained from the Mayo Clinic EHR as well as the MOA-2 gold standard. Since we have a limited number of women with bilateral oophorectomy only, we reported both macro average metrics for overall surgery status (which calculated the matrix independently by surgery status but not considering weights for sample size) and weighted average metrics for overall surgery status (with weighting by sample size). Recognizing that missing data are common owing to movement in and out of health care systems, we also analyzed the recovery ratio of the surgery status information (using the weighted average F1-score) between the limited and reverse-chronological years of records and the total years of records by the NLP algorithm.

Discrepancy Analysis

After training and validating the NLP algorithm, it was subsequently applied to all remaining Mayo records. All discrepancies between NLP classifications and gold standard MOA-2 data were then identified and manually reviewed by 1 annotator (EDM), which were subsequently classified into 3 categories: external information gaps, internal information gaps, and technical errors by the NLP algorithm. External discrepancies were defined as differences in surgical status between the 2 sources (eg, the gold standard categorizes a woman as having surgery, but the surgery is not mentioned in the Mayo Clinic EHR) and were reviewed by another annotator (LGR, a physician) to determine the true surgical status. Internal discrepancies were differences due to inconsistent or inaccurate surgery history information in the Mayo Clinic EHR (eg, partial vs complete surgery). Finally, we also identified technical errors by the NLP algorithm (eg, negated but classified as positive).

Ethics Approval

The study was approved by the Mayo Clinic and Olmsted Medical Center Institutional Review Boards.

Results

Corpus Analysis and Results of the NLP Algorithm on the Train and Test Sets

In this cohort, the median age at follow-up was 60 years (IQR 54-66 years), and the median length of follow-up was 16.2 years (IQR 11.1-21.1 years). Among 3870 women, 1473 (38.1%) did not undergo gynecologic surgery, while 2397 (61.9%) underwent at least one gynecologic surgery before their latest follow-up date. Most women with gynecologic surgery history (2069/2397, 86.3%) had only 1 surgical date, 12.7% (304/2397) had 2 separate surgery dates, and 1% (24/2397) had 3 separate surgery dates.

Among the 54 cases selected for interannotator reliability assessment, the percentage agreement was 90.7% (49/54) and the kappa statistic was 0.85. Of the 530 patients initially selected for annotation, 446 (84.2%) were accurately annotated using the Mayo EHR compared to the MOA-2 gold standard (Multimedia Appendix 3). In general, disagreement between Mayo-annotated and MOA-2 gynecologic surgery statuses was large with respect to false negatives (ie, Mayo annotations inaccurately assigned to “no surgery”), which comprised 59 of the 84 total misclassifications (70.2%).

We present the test-set performance metrics relative to the Mayo annotation labels and MOA-2 gold standard labels in Table 2 (and train set performances reported in Multimedia Appendix 4). Using surgical statuses extracted from the Mayo EHR, the NLP algorithm correctly classified 82.3% of women (218/265 women in the test set), with weighted averages of 0.85 precision, 0.82 recall, and 0.83 F1-score. When compared to the MOA-2 labels, performance dropped moderately and the surgical status

of 76.2% of the women (202/265 women in the test set) was correctly classified by the NLP algorithm through the follow-up date. The NLP algorithm achieved precision of 0.79, recall of 0.76, and weighted average F1-score of 0.76 in the test set (Table 2). Performance measures varied by the surgery type, with the lowest performance observed for assessing "bilateral oophorectomy only" and the highest for identifying "hysterectomy only."

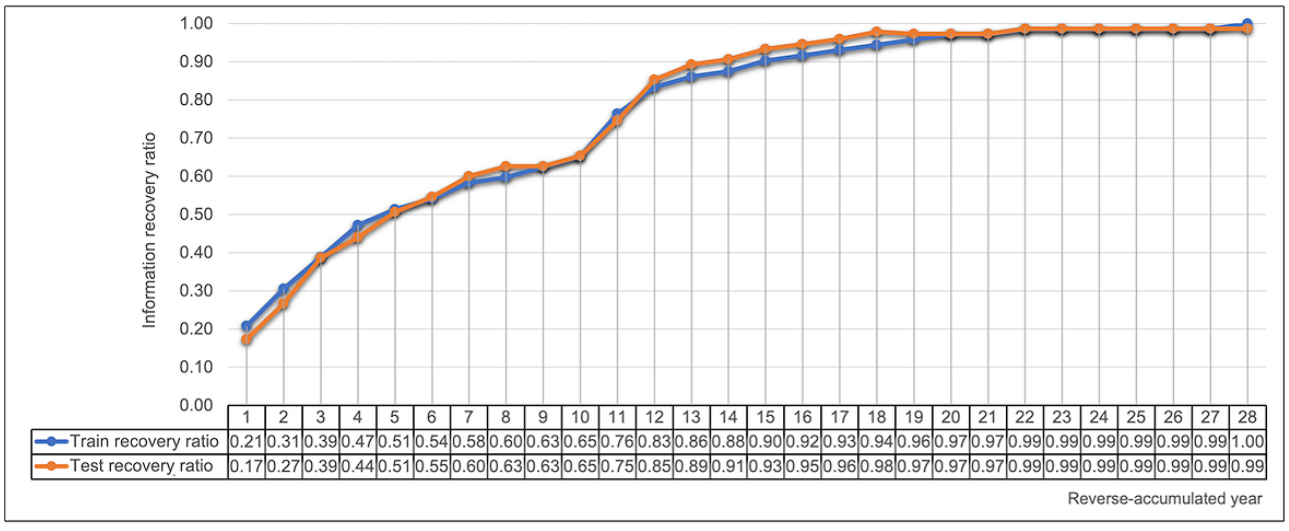
Table 2. Test set evaluation (n=265) of the natural language processing algorithm using Mayo and Mayo Clinic Cohort Study of Oophorectomy and Aging-2 annotations.

Algorithm surgery type	Mayo			MOA-2[a]		
	Precision	Recall	F1-score	Precision	Recall	F1-score
No surgery	0.98	0.81	0.89	0.74	0.96	0.83
Bilateral oophorectomy only	0.42	0.62	0.50	0.58	0.47	0.52
Hysterectomy and bilateral oophorectomy	0.68	1.00	0.81	0.62	0.90	0.73
Unilateral oophorectomy only	0.94	0.74	0.83	0.97	0.66	0.79
Hysterectomy and unilateral oophorectomy	0.84	0.71	0.77	0.84	0.64	0.73
Hysterectomy only	0.81	0.88	0.84	0.86	0.74	0.80
Overall macro average	0.78	0.79	0.77	0.77	0.73	0.73
Overall weighted average	0.85	0.82	0.83	0.79	0.76	0.76

[a]MOA-2: Mayo Clinic Cohort Study of Oophorectomy and Aging-2.

If we restricted the NLP algorithm to use recent clinic notes in the reverse-chronological order from individual follow-up dates, 1 year of clinical notes yielded only 17.1% (0.13/0.76) of the surgical status information compared to the original weighted average F1-score of 0.76. A minimum of 14 years of narrative notes in the test set was required for the NLP algorithm to recover 90% of the surgical status information. The overall trend of the weighted average F1-score recovery ratio according to reverse-chronological year is represented in Figure 2. About 62.3% (268/430) of women had the surgery date information present in at least one clinical note. We also observed a disparity in date information by surgery status. Specifically, only 23% (46/200) of women with unilateral oophorectomy surgery had the date information present. In contrast, 70% (91/130) of women with bilateral oophorectomy and 82.7% (248/300) of women with hysterectomy had the date information present.

Figure 2. Recovery ratio for the surgery status information by years of electronic health record data available.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Train recovery ratio	0.21	0.31	0.39	0.47	0.51	0.54	0.58	0.60	0.63	0.65	0.76	0.83	0.86	0.88	0.90	0.92	0.93	0.94	0.96	0.97	0.97	0.99	0.99	0.99	0.99	0.99	0.99	1.00
Test recovery ratio	0.17	0.27	0.39	0.44	0.51	0.55	0.60	0.63	0.63	0.65	0.75	0.85	0.89	0.91	0.93	0.95	0.96	0.98	0.97	0.97	0.97	0.99	0.99	0.99	0.99	0.99	0.99	0.99

Results of the NLP Algorithm on the Remaining Set and Discrepancy Analysis

When we applied the NLP algorithm to the remaining set (n=3340), we correctly classified 86.1% (2876/3340) of the surgery status of patients. Similar to the test set results, recall rates were relatively poor for positive surgical history. In Table 3, we summarized 464 discrepancies of surgical status in the NLP algorithm classification compared to the multi-institutional MOA-2 gold standard. First, 6% (199/3340) of women in this set had either no recorded surgery information or partial information in the EHR. Second, we found inconsistent or inaccurate information for 4.3% (144/3340) of women. Lastly, the NLP algorithm misclassified the surgery status of 3.6%

(121/3340) of women. External information gaps represented 42.9% (199/464) of the discrepancies, internal information gaps represented 31% (144/464) of the discrepancies, and 26.1% (121/464) were technical errors of the NLP algorithm.

Table 3. Summary of the 464 discrepancies observed.

Type, categorization	Value (n)
External information gap	
Mayo electronic health record	
Missing information	92
Partial information	49
Gold standard	
Missing information	11
Partial information	45
Both	
Partial information	2
Internal information gap	
Correction over time	
Documented surgeries but revealed later as no surgeries	74
Irregular concept scope	
Partial surgery versus complete surgery	17
Biopsy examination versus complete surgery	23
Planned surgery versus real surgery	5
Miscommunication within clinical documents	
Hysterectomy versus hysteroscopy	8
Incorrect laterality (left vs right-side) information	12
Typo	5
Technical errors of natural language processing	
Complicated context	
Discussion versus real surgery	63
Family history versus history of patient	11
Complex expressions of partial surgery	6
Complex expressions of laterality information	12
Incorrect certainty	
Negated but classified as positive	10
Positive but classified as negated	1
Positive but classified as hypothetical	1
Unknown features	
Irrelevant section header	2
Unknown keywords/rules	15

Of the 199 external information gaps, positive surgical history was missing in the Mayo Clinic EHR for 92 women (ie, false negatives, 46.2%). In contrast, the surgical history present in the Mayo Clinic EHR was not captured by the gold standard for 11 women. There were discrepancies related to surgery type for 96 women. The details for all external information gaps are summarized in Multimedia Appendix 5.

Of the 144 internal information gaps identified, the chart review revealed multiple potential sources of inconsistency. The details regarding the surgery type were frequently inconsistent, and about half of the discrepancies (n=74) resulted from the correction of surgery information over time. For example, one note for a patient indicated “BSO” (bilateral salpingo-oophorectomy), whereas all other notes contained “remained right ovary,” indicating a unilateral oophorectomy.

There were differences between clinical notes and the more detailed surgical or pathology reports (n=45). Miscommunication within clinical documents in the use of words (eg, misinterpretation of "hysteroscopy" noted as "hysterectomy"), incorrect laterality (ie, left vs right), and typos were also observed (n=25).

Finally, there were 121 technical errors in the NLP algorithm. The NLP algorithm had difficulties in accurately processing complicated contextual information (n=92). For example, it had difficulty distinguishing discussion/consideration from real surgery or patient history from family history. In addition, the NLP algorithm misclassified certainty information of sentences (eg, negated but classified as positive, n=12), or it missed the surgical information owing to the limited set of keywords/phrases or associated section header information (n=17). For example, a subtitle in the surgery operative notes, "Uterus, endometrium, hysterectomy: Inactive" was classified as a valid "hysterectomy" by the NLP algorithm.

Discussion

A comprehensive medical history of individual patients is necessary to achieve a high quality of patient care and to support clinical research. Identifying historical surgery information is challenging because some surgeries may have occurred decades before the widespread adoption of EHR systems. Furthermore, useful information is often distributed in separate EHR systems owing to the preference or needs of the patients. Finally, limited time during clinic visits and quality of self-reported history often result in incomprehensive surgery information. This study sought to extract gynecological surgical history from a single EHR by using a rule-based NLP algorithm and to compare these results with gold standard data ascertained from a manual multi-institutional record review.

The NLP algorithm that was trained on surgery statuses manually extracted from the Mayo EHR was largely successful with respect to being internally valid; however, false negatives were commonly encountered when compared to gold standard information. In addition to misclassification, the date of the surgery was often missing, rendering ascertainment of surgery timing difficult. The preponderance of false negatives is consistent with a model of information loss over the lifetime of a patient and may serve as a source of systematic bias in research.

The external information gaps were the most common errors encountered and related to missing or incomplete information in the EHR for surgery status or surgery type. Similar to the test set results, we observed that nearly 50% of the external discrepancies were false negative in nature. These results starkly contrast diagnostic code–based results reported by Rocca et al [16] when using the full resources of the REP to build the MOA-2 cohort, which were highly accurate in identifying surgical history status for oophorectomy. In addition, the longer a woman was followed in the EHR, the more likely her gynecologic surgery was recorded in the clinical narrative. This is again fairly intuitive, as follow-up time within a single EHR system likely captures consistent and reliable information with fewer opportunities for data loss in record transfers. Surgical date information was sparse and differed by surgery type. For example, patients commonly provided their age at the time of surgery rather than the surgery date. Consequently, research that relies on reliable ascertainment of surgery dates should take these heterogeneous and complex modes of information representation into consideration. The most common internal information gap identified was inaccurate reporting of surgical status by the clinician, the patient, or both that was subsequently refuted. Thus, information conflict resolution is another critical element to address in information extraction from long-term clinical narratives.

With the growing popularity of utilizing NLP-based phenotyping for research using EHR data, it is important to consider the nonnegligible risk of misclassification despite evidence of internal validity for NLP-based phenotyping algorithms. Systematic misclassification toward false negatives could induce biases in research, particularly for patient populations that are highly transient and may change care providers frequently. Strategies to reduce information gaps and to improve the collection of surgical history include leveraging the NLP technology with optical character recognition technology to digitalize paper-based records or acquiring the records digitally via a health information exchange [22-25]. Lastly, the implementation of systematic questionnaires to gather prior surgical information may significantly reduce information gaps as well. The questionnaires can also be leveraged for capturing potential documentation errors besides enhancing documentation quality.

The strengths of our study include the total sample size available and the high-quality gold standard phenotype data. However, the performance of the simple rule-based NLP algorithm could be improved upon with more sophisticated methods, as indicated by the extent of technical errors identified in the discrepancy analysis.

In conclusion, our study demonstrated the overall feasibility of extracting gynecological surgeries that often predated the EHR system by decades using a rule-based NLP algorithm. However, we identified external and internal information gaps by comparing NLP algorithm results to a manually abstracted gold standard. Additional efforts are necessary to mitigate these information gaps and include the use of advanced NLP techniques to process paper medical records and systematic collection and documentation of surgical history.

Acknowledgments

We would like to thank the Mayo Clinic Cohort Study for Oophorectomy and Aging-2 team (CYS, WAR, LGR) for providing the gold standard cohort and Ellen E Koepsell for conducting annotations. We also thank Liwei Wang, Sheila M Manemann, Jennifer L St. Sauver, Paul Y Takahashi, Janet E Olson, Virginia M Miller, Véronique L Roger, Paul A Decker, Jill M Killian,

and Mary G Roberts for providing insightful comments. This work was supported by National Institutes of Health grants HL136659 and U01TR002062. This study was made possible using the resources of the Rochester Epidemiology Project, which is supported by the National Institute on Aging of the National Institutes of Health under awards AG034676 and AG052425.

Authors' Contributions

SM designed the methods and experiments, generated a data set, implemented the algorithm, analyzed the data, and interpreted the results. EDM and LGR conducted annotation and validation of the algorithm. SM carried out the experiments. SM drafted the manuscript. LAC and BSAK helped to draft the manuscript. SJB, HL, and NBL conceived the study and helped to draft the manuscript. All authors read and approved the manuscript.

Conflicts of Interest

WAR received research support from the National Institutes of Health (R01 AG052425, RF1 AG055151, R33 AG058738, and U54 AG044170).

Multimedia Appendix 1
Commonly used keywords and phrases to describe gynecologic surgeries and associated clinical note section headers used in the algorithm.
[DOCX File , 14 KB - jmir_v24i1e29015_app1.docx]

Multimedia Appendix 2
Rule to determine final gynecologic surgery status.
[DOCX File , 15 KB - jmir_v24i1e29015_app2.docx]

Multimedia Appendix 3
Annotation comparison between Mayo and Mayo Clinic Cohort Study of Oophorectomy and Aging-2 gold standard.
[DOCX File , 15 KB - jmir_v24i1e29015_app3.docx]

Multimedia Appendix 4
Training set (n=265) evaluation of the natural language processing algorithm using Mayo and Mayo Clinic Cohort Study of Oophorectomy and Aging-2 annotations.
[DOCX File , 15 KB - jmir_v24i1e29015_app4.docx]

Multimedia Appendix 5
Distribution of the 199 external information discrepancies observed.
[DOCX File , 14 KB - jmir_v24i1e29015_app5.docx]

References

1. Parker WH, Broder MS, Chang E, Feskanich D, Farquhar C, Liu Z, et al. Ovarian conservation at the time of hysterectomy and long-term health outcomes in the nurses' health study. Obstet Gynecol 2009 May;113(5):1027-1037 [FREE Full text] [doi: 10.1097/AOG.0b013e3181a11c64] [Medline: 19384117]
2. Parker WH, Feskanich D, Broder MS, Chang E, Shoupe D, Farquhar CM, et al. Long-term mortality associated with oophorectomy compared with ovarian conservation in the nurses' health study. Obstet Gynecol 2013 Apr;121(4):709-716 [FREE Full text] [doi: 10.1097/AOG.0b013e3182864350] [Medline: 23635669]
3. Rocca WA, Grossardt BR, de Andrade M, Malkasian GD, Melton LJ. Survival patterns after oophorectomy in premenopausal women: a population-based cohort study. Lancet Oncol 2006 Oct;7(10):821-828. [doi: 10.1016/S1470-2045(06)70869-5] [Medline: 17012044]
4. Melton LJ, Crowson CS, Malkasian GD, O'Fallon WM. Fracture risk following bilateral oophorectomy. J Clin Epidemiol 1996 Oct;49(10):1111-1115. [doi: 10.1016/0895-4356(96)00211-9] [Medline: 8826990]
5. Rocca WA, Gazzuola-Rocca L, Smith CY, Grossardt BR, Faubion SS, Shuster LT, et al. Accelerated Accumulation of Multimorbidity After Bilateral Oophorectomy: A Population-Based Cohort Study. Mayo Clin Proc 2016 Nov;91(11):1577-1589 [FREE Full text] [doi: 10.1016/j.mayocp.2016.08.002] [Medline: 27693001]
6. Rocca WA, Gazzuola Rocca L, Smith CY, Grossardt BR, Faubion SS, Shuster LT, et al. Loss of Ovarian Hormones and Accelerated Somatic and Mental Aging. Physiology (Bethesda) 2018 Nov 01;33(6):374-383 [FREE Full text] [doi: 10.1152/physiol.00024.2018] [Medline: 30303778]
7. Gandhi TK, Sittig DF, Franklin M, Sussman AJ, Fairchild DG, Bates DW. Communication breakdown in the outpatient referral process. J Gen Intern Med 2000 Sep;15(9):626-631 [FREE Full text] [doi: 10.1046/j.1525-1497.2000.91119.x] [Medline: 11029676]

8. Zuchowski JL, Chrystal JG, Hamilton AB, Patton EW, Zephyrin LC, Yano EM, et al. Coordinating Care Across Health Care Systems for Veterans With Gynecologic Malignancies: A Qualitative Analysis. Med Care 2017 Jul;55 Suppl 7 Suppl 1:S53-S60. [doi: 10.1097/MLR.0000000000000737] [Medline: 28614184]
9. Mårtensson L, Hensing G. Health literacy -- a heterogeneous phenomenon: a literature review. Scand J Caring Sci 2012 Mar;26(1):151-160. [doi: 10.1111/j.1471-6712.2011.00900.x] [Medline: 21627673]
10. Devraj R, Herndon CM, Griffin J. Pain awareness and medication knowledge: a health literacy evaluation. J Pain Palliat Care Pharmacother 2013 Mar;27(1):19-27. [doi: 10.3109/15360288.2012.751955] [Medline: 23379354]
11. Carayon P, Alyousef B, Hoonakker P, Hundt AS, Cartmill R, Tomcavage J, et al. Challenges to care coordination posed by the use of multiple health IT applications. Work 2012;41 Suppl 1:4468-4473. [doi: 10.3233/WOR-2012-0746-4468] [Medline: 22317409]
12. Dolezel D, Moczygemba J. Implementing EHRs: An Exploratory Study to Examine Current Practices in Migrating Physician Practice. Perspect Health Inf Manag 2015;12:1e [FREE Full text] [Medline: 26807077]
13. Penrod LE. Electronic Health Record Transition Considerations. PM R 2017 May;9(5S):S13-S18. [doi: 10.1016/j.pmrj.2017.01.009] [Medline: 28527498]
14. Victoor A, Delnoij DMJ, Friele RD, Rademakers JJDJM. Determinants of patient choice of healthcare providers: a scoping review. BMC Health Serv Res 2012 Aug 22;12:272 [FREE Full text] [doi: 10.1186/1472-6963-12-272] [Medline: 22913549]
15. Rocca WA, Gazzuola Rocca L, Smith CY, Grossardt BR, Faubion SS, Shuster LT, et al. Bilateral Oophorectomy and Accelerated Aging: Cause or Effect? J Gerontol A Biol Sci Med Sci 2017 Sep 01;72(9):1213-1217 [FREE Full text] [doi: 10.1093/gerona/glx026] [Medline: 28329133]
16. Rocca WA, Gazzuola Rocca L, Smith CY, Grossardt BR, Faubion SS, Shuster LT, et al. Cohort profile: the Mayo Clinic Cohort Study of Oophorectomy and Aging-2 (MOA-2) in Olmsted County, Minnesota (USA). BMJ Open 2017 Nov 20;7(11):e018861 [FREE Full text] [doi: 10.1136/bmjopen-2017-018861] [Medline: 29162577]
17. Rocca WA, Yawn BP, St Sauver JL, Grossardt BR, Melton LJ. History of the Rochester Epidemiology Project: half a century of medical records linkage in a US population. Mayo Clin Proc 2012 Dec;87(12):1202-1213 [FREE Full text] [doi: 10.1016/j.mayocp.2012.08.012] [Medline: 23199802]
18. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ, Rocca WA. Use of a medical records linkage system to enumerate a dynamic population over time: the Rochester epidemiology project. Am J Epidemiol 2011 May 01;173(9):1059-1068 [FREE Full text] [doi: 10.1093/aje/kwq482] [Medline: 21430193]
19. Rocca WA, Grossardt BR, Brue SM, Bock-Goodner CM, Chamberlain AM, Wilson PM, et al. Data Resource Profile: Expansion of the Rochester Epidemiology Project medical records-linkage system (E-REP). Int J Epidemiol 2018 Apr 01;47(2):368-368j [FREE Full text] [doi: 10.1093/ije/dyx268] [Medline: 29346555]
20. St Sauver JL, Grossardt BR, Yawn BP, Melton LJ, Pankratz JJ, Brue SM, et al. Data resource profile: the Rochester Epidemiology Project (REP) medical records-linkage system. Int J Epidemiol 2012 Dec;41(6):1614-1624 [FREE Full text] [doi: 10.1093/ije/dys195] [Medline: 23159830]
21. Wen A, Fu S, Moon S, El Wazir M, Rosenbaum A, Kaggal VC, et al. Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation. NPJ Digit Med 2019;2:130 [FREE Full text] [doi: 10.1038/s41746-019-0208-8] [Medline: 31872069]
22. Moon S, Liu S, Chen D, Wang Y, Wood DL, Chaudhry R, et al. Salience of Medical Concepts of Inside Clinical Texts and Outside Medical Records for Referred Cardiovascular Patients. J Healthc Inform Res 2019 Jan 28;3(2):200-219. [doi: 10.1007/s41666-019-00044-5]
23. Li X, Hu G, Teng X, Xie G. Building structured personal health records from photographs of printed medical records. 2015 Presented at: AMIA Annu Symp Proc; November 5; San Francisco p. 833-842.
24. Vest JR, Zhao H, Jasperson J, Gamm LD, Ohsfeldt RL. Factors motivating and affecting health information exchange usage. J Am Med Inform Assoc 2011;18(2):143-149 [FREE Full text] [doi: 10.1136/jamia.2010.004812] [Medline: 21262919]
25. Furukawa MF, King J, Patel V, Hsiao C, Adler-Milstein J, Jha AK. Despite substantial progress In EHR adoption, health information exchange and patient engagement remain low in office settings. Health Aff (Millwood) 2014 Sep;33(9):1672-1679. [doi: 10.1377/hlthaff.2014.0445] [Medline: 25104827]

Abbreviations

EHR: electronic health record
MOA-2: Mayo Clinic Cohort Study for Oophorectomy and Aging-2
NLP: natural language processing
REP: Rochester Epidemiology Project

Edited by A Mavragani; submitted 24.03.21; peer-reviewed by Y An, L Ferreira, M Antoniou; comments to author 18.05.21; revised version received 13.07.21; accepted 01.12.21; published 28.01.22.

Please cite as:
Moon S, Carlson LA, Moser ED, Agnikula Kshatriya BS, Smith CY, Rocca WA, Gazzuola Rocca L, Bielinski SJ, Liu H, Larson NB Identifying Information Gaps in Electronic Health Records by Using Natural Language Processing: Gynecologic Surgery History Identification
J Med Internet Res 2022;24(1):e29015
URL: https://www.jmir.org/2022/1/e29015
doi:10.2196/29015
PMID:35089141

Original Paper

Adherence, Efficacy, and Safety of Wearable Technology–Assisted Combined Home-Based Exercise in Chinese Patients With Ankylosing Spondylitis: Randomized Pilot Controlled Clinical Trial

Yiwen Wang[1*], PhD; Xingkang Liu[1,2*], PhD; Weimin Wang[3*], MD; Yanyun Shi[3], MD; Xiaojian Ji[1], PhD; Lidong Hu[1], PhD; Lei Wang[1], PhD; Yiquan Yin[4], PhD; Siyuan Xie[2], MPT; Jian Zhu[1], Prof Dr; Jianglin Zhang[1], Prof Dr; Wei Jiao[2], Prof Dr; Feng Huang[1,5], Prof Dr

[1]Department of Rheumatology and Immunology, First Medical Center, Chinese People's Liberation Army General Hospital, Beijing, China

[2]College of Sport Medicine and Sport Rehabilitation, Beijing Sport University, Beijing, China

[3]Health Management Institute, Second Medical Center, Chinese People's Liberation Army General Hospital, Beijing, China

[4]Advanced Research Institute of Olympic, Beijing Sport University, Beijing, China

[5]National Clinical Research Center for Orthopedics, Sports Medicine and Rehabilitation, Chinese People's Liberation Army General Hospital, Beijing, China

[*]these authors contributed equally

Corresponding Author:
Feng Huang, Prof Dr
Department of Rheumatology and Immunology
First Medical Center
Chinese People's Liberation Army General Hospital
28 Fuxing Road
Beijing, 100853
China
Phone: 86 010 5549 9314
Email: fhuang@301hospital.com.cn

Abstract

Background: Clinical practice guidelines recommend that exercise is essential in the self-management of ankylosing spondylitis (AS). Attending supervised interventions requiring periodic medical center visits can be difficult and patients may decline participation, whereas effective home-based exercise interventions that do not require regular medical center visits are likely to be more accessible for AS patients.

Objective: The goal of the research was to investigate the adherence, efficacy, and safety of a wearable technology–assisted combined home-based exercise program in AS.

Methods: This was a 16-week investigator-initiated, assessor-blinded, randomized, pilot controlled trial conducted at Chinese People's Liberation Army General Hospital. We enrolled patients with AS who had no regular exercise habits and had been stable in drug treatment for the preceding month. Patients were randomly assigned (1:1) using a computer algorithm. An exercise program consisting of moderate-intensity aerobic exercise and functional exercise was given to the patients in the intervention group. The exercise intensity was controlled by a Mio FUSE Heart Rate Monitor wristband, which uses photoplethysmography to measure heart rate. Patients in the control group received usual care. The primary outcome was the difference in the Ankylosing Spondylitis Disease Activity Score (ASDAS). The secondary outcomes were patient global assessment (PGA), physician global assessment (PhGA), total pain, nocturnal pain, Bath Ankylosing Spondylitis Disease Activity Index (BASDAI), BAS Functional Index (BASFI), BAS Metrology Index (BASMI), Spondyloarthritis International Society Health Index (ASAS HI), 36-item Short Form Survey (SF-36), maximal oxygen uptake (VO_2) max, body composition, range of motion of joints, and muscle endurance tests. Retention rate, adherence rate, barriers to being active, and adverse events were also assessed.

Results: A total of 77 patients were screened, of whom 55 (71%) patients were enrolled; 2% (1/55) withdrew without treatment after randomization. Patients were assigned to the intervention (n=26) or control group (n=28). The median adherence rate of the prescribed exercise protocol was 84.2% (IQR 48.7%-97.9%). For the primary outcome, between-group difference of ASDAS was significant, favoring the intervention (–0.2, 95% CI –0.4 to 0.02, P=.03). For the secondary outcomes, significant between-group

differences at 16 weeks were detected in PGA, PhGA, total pain, BASDAI, BASDAI-fatigue, BASDAI–spinal pain, BASDAI–morning stiffness intensity, BASFI, and BASMI. Moreover, the frequency of difficulty in ASAS HI-motivation at 16 weeks was less in the intervention group (P=.03). Between-group difference for change from baseline were also detected in VO_2 max, SF-36, back extensor endurance test, and the range of motion of cervical lateral flexion at 16 weeks. Lack of time, energy, and willpower were the most distinct barriers to being active. Incidences of adverse events were similar between groups (P=.11).

Conclusions: Our pilot study suggests that this technology-assisted combined home-based exercise program can improve the clinical outcomes of patients with AS who have no exercise habit, with good adherence and safety profile.

Trial Registration: Chinese Clinical Trial Registry ChiCTR1900024244; http://www.chictr.org.cn/showproj.aspx?proj=40176

(J Med Internet Res 2022;24(1):e29703) doi:10.2196/29703

KEYWORDS

ankylosing spondylitis; wearable technology; home-based exercise; combined exercise; randomized controlled trial; RCT; exercise; wearable; photoplethysmography; spondyloarthritis

Introduction

Ankylosing spondylitis (AS), the prototype of spondyloarthritis (SpA), is an inflammatory disease that can affect the axial skeleton and peripheral joints [1]. Since AS usually starts in young adulthood, the structural and functional impairments resulting in limitation of activities and social participation can exert considerable lifetime individual impact [2].

Clinical practice guidelines for AS recommend that exercise be included in the management of AS [3,4]. The American College of Sports Medicine (ACSM) recommends that all healthy adults participate in moderate intensity aerobic physical activity (PA) for a minimum of 30 minutes 5 days per week or vigorous intensity aerobic activity for a minimum of 20 minutes 3 days per week [5]. In 2018, a European League Against Rheumatism (EULAR) task force indicated that promoting PA consistent with general PA recommendations and conducting combined PA in the 4 domains (cardiorespiratory fitness, muscle strength, flexibility, and neuromotor performance) should be an integral part of standard care of inflammatory arthritis and osteoarthritis [6].

Although supervised exercise has advantages, it is difficult for most patients to access physiotherapists specializing in AS, and attending supervised interventions requiring periodic medical center visits can be difficult and expensive [7,8]. Conversely, home-based exercise interventions that do not require regular medical center visits are likely to be cheaper and more accessible. Currently, poor adherence and lack of monitoring strategy are 2 barriers to improving and maintaining the quality of exercise interventions [9].

Digital health encompassing a broad array of technologies has the potential to improve the quality of musculoskeletal disease care [10]. Wearable technology, one of the digital health technologies based on a wearable activity tracker (WAT), is considered to be effective to deliver exercise intervention [10]. EULAR recommendations indicate that self-monitoring tools including WAT are identified for PA assessments [6,11]. In addition, WAT can provide effective behavioral change techniques to facilitate patients to acquire skills in self-monitoring, goal setting, action planning, and feedback and problem solving in home-based exercise [6]. Currently, most evidence about WAT is for rheumatoid arthritis, osteoarthritis, and juvenile idiopathic arthritis; evidence concerning AS is scarce [10,12-15]. Therefore, we conducted this clinical trial to investigate the adherence, efficacy, and safety of this wearable technology-assisted combined home-based exercise intervention in patients with AS.

Methods

Study Design

This was a 16-week, randomized, open-label, accessor-blinded, controlled clinical trial conducted at the Chinese People's Liberation Army (PLA) General Hospital (Multimedia Appendix 1). Enrolled patients were allocated with a 1:1 ratio and assessed at baseline, 8 weeks, and 16 weeks by trained research staff blinded to group assignment. The study was approved by the ethics committee at the Chinese PLA General Hospital (S2019-118-01), conducted in compliance with the Helsinki Agreement, and registered with the Chinese Clinical Trial Registry [ChiCTR 1900024244]. Informed consent was obtained from all eligible participants.

Identification of Potential Participants

Potential participants were identified using a passive online recruitment approach through the Smartphone SpondyloArthritis Management System (SpAMS), which was created to provide patient education and deliver advice on disease management in patients with AS in China [16]. SpAMS was linked to WeChat (first released in 2011; Tencent Holdings Ltd), an instant messaging social network in China, and can be leveraged for professional purposes. Details have been published elsewhere [16,17]. As volunteers were identified, inclusion criteria were initially confirmed during online screening interviews. Written confirmation of AS diagnosis was required from the participant's board-certified rheumatologist. If volunteers met the inclusion criteria, face-to-face interviews were conducted at the clinic for final confirmation of all inclusion criteria. Participants meeting all inclusion criteria were enrolled consecutively.

Inclusion and Exclusion Criteria

Inclusion criteria were fulfillment of the criteria for AS (1984 Modified New York criteria) [18], aged 18 to 60 years, stable drug treatment in the preceding month, and Ankylosing Spondylitis Disease Activity Score (ASDAS) between 1.3 and 3.5. Exclusion criteria were cardiovascular disease or clinical

status at high risk, screened with the American Heart Association/ACSM Health/Fitness Facility Preparticipation Screening Questionnaire [19], cervical vertebral bridges, surgery within the preceding 6 months, biological agents (tumor necrosis factor inhibitor therapy, etc) used in the preceding 3 months, regular exercise in the preceding 3 months (eg, yoga, Tai Chi, Baduanjin 3 or more times per week, 20 minutes per time), and factors leading to the inability to receive regular exercise rehabilitation (such as language impairment, difficulty in understanding, and limited movements).

Randomization and Masking

Once the informed consent was signed and baseline assessments were completed, patients were randomly allocated to the intervention or control arm with 1:1 allocation ratio using a computer-generated randomization list performed by a research nurse not associated with the intervention portion of the study. The assessment staff and statisticians were masked to the group assignment.

Interventions

A 16-week combined exercise program consisting of in-person counseling sessions, supervised training sessions, and aerobic and functional home-based exercise was given to patients in the intervention group after randomization.

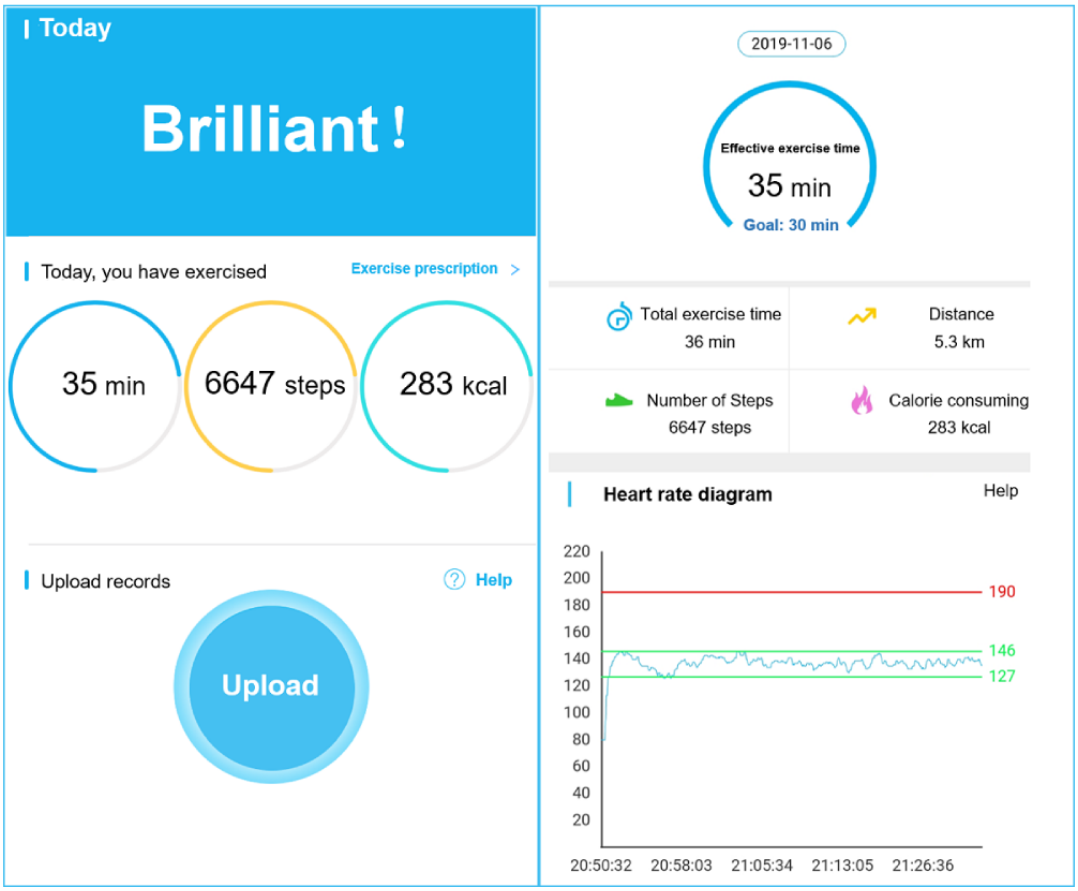
In-Person Counseling Session and Brief Supervised Training Sessions

At baseline, in-person counseling sessions were held by trained research staff. The in-person counseling session is a structured interview containing the 4 domains: health benefits of exercise; overview of this exercise program; bullet points on effective and safe exercise; and how to use the wearable devices in this exercise program. In addition, supervised training sessions including a 30-minute aerobic exercise and 60 minutes of functional exercise were given for 2 consecutive days by a physiotherapist at baseline and 8 weeks to each patient assigned to the intervention group.

Aerobic Exercise

Aerobic exercise at a moderate intensity of 64% to 76% maximal heart rate (HRmax) was prescribed, and the exercise type was brisk walking or running (Multimedia Appendix 2) [5]. During each session, the exercise intensity was monitored and controlled by a Mio FUSE Heart Rate Monitor wristband (Medisana GmbH), which uses photoplethysmography to measure heart rate. The wristband was synchronized with the smartphone app (G health, version 2.7.1) via Bluetooth. Exercise was considered effective only when it reached moderate intensity. The prescribed protocol was 30 minutes of effective aerobic exercise 5 days per week. Data including the duration of effective aerobic exercise and heart rate during each session was all uploaded to the cloud virtual machine and the smartphone app (Figure 1).

Figure 1. Screenshots from the treatment app. Text has been translated from Chinese to English for illustration purposes.



Functional Exercise

Functional exercise consisting of posture, range of motion, resistance, stability, and stretching exercises was prescribed for 60 minutes 3 days per week (Multimedia Appendix 2). Individualized written functional exercise plans including key points of each movement were given to patients to help them perform home-based functional exercise and revised at the 8-week visit.

Usual Care

Both groups were permitted to receive routine medical therapy. Medical therapy included nonsteroidal anti-inflammatory drugs, conventional disease-modifying antirheumatic drugs including methotrexate and sulfasalazine. Biological agent or new conventional disease-modifying antirheumatic drugs were not allowed during the trial phase. Patients in the control group were asked to maintain usual PA level during the 16-week follow-up.

Monitoring and Adherence

The sessions of aerobic exercise were recorded and analyzed based on the effective duration of exercise but not the total time spent on exercise with the consideration of better controlling the exercise intensity and duration. The duration of effective exercise was presented in the interface of the smartphone app, and the number of sessions fulfilling the exercise target was uploaded to the cloud. Adherence to the functional exercise plan was registered by uploading pictures taken after each session by the patient. At the end of the trial, the adherence rate was calculated with number of the sessions attended divided by number of sessions prescribed (128 sessions including aerobic and functional exercise sessions in 16 weeks). In addition, the Barriers to Being Active Quiz from the Centers for Disease Control and Prevention was completed by patients who participated in less than 80% of the prescribed exercise protocol in the intervention group [20].

Outcome Assessments

An assessment was conducted at baseline (before randomization), 8 weeks, and 16 weeks by trained research staff blinded to group assignment including demographic information and primary and secondary outcomes. Investigators were blinded to outcome data until the end of the trial.

Primary Outcome

The primary outcome was the between-group difference for change from baseline to 16 weeks in the ASDAS [21]. ASDAS includes total back pain, patient global assessment of disease activity, peripheral pain/swelling, duration of morning stiffness, and C-reactive protein level.

Secondary Outcomes

In addition to the ASDAS, the Bath Ankylosing Spondylitis Disease Activity Index (BASDAI) was used to reflect disease activity. Moreover, the patient global assessment (PGA), physician global assessment (PhGA), spinal pain–total pain, spinal pain–nocturnal pain, Bath Ankylosing Spondylitis Functional Index (BASFI) [22], and Bath Ankylosing Spondylitis Metrology Index (BASMI) [23] were evaluated. Assessment using the Spondyloarthritis International Society Health Index (ASAS HI) was performed to determine the common difficulties of patients with AS [24]. Additionally, health-related quality of life was measured using the 36-Item Short Form Survey (SF-36) including physical function, role–physical, bodily pain, general health, vitality, social functioning, role–emotional, and mental health [25].

Cardiorespiratory fitness was tested with submaximal exercise tests (multistage model) on a treadmill. Based on the workload at the end of the test and estimated HRmax, the maximal oxygen uptake (VO_2 max) was estimated [26].

Body composition including lean body mass, percentage of body fat, and visceral fat area was evaluated using the noninvasive bioelectrical impedance analysis method with the InBody 770 Body Analyzer (InBody Co, LTD).

The Clinometer smartphone app (version 3.7, Plaincode Software Solutions) was used to determine the range of motion of the cervical spine and hip joints. Back extensor and flexor endurance tests were conducted to evaluate lumbar muscle endurance.

Sample Size

The study was designed with a planned sample size of 54 patients with a 1:1 group allocation ratio. Determination of the sample size was based on detecting a medium effect size of 0.25 calculated through a standardized mean difference of the change in ASDAS from baseline to 16 weeks between groups. Calculations were performed with G*Power (version 3.1, Heinrich Heine University) software. With a power of 95% or higher to detect differences between groups, 22 patients were calculated to be allocated to each group with a type I error rate of 5%. The loss to follow-up rate was assumed to be 20%. Therefore, the sample size of this trial was determined to be 27 patients in each group.

Statistical Analysis

Primary and secondary outcomes were analyzed according to intention-to-treat principles by including all patients who were randomly allocated to either group and underwent at least 1 efficacy assessment. Last observation carried forward was used for missing observations. Separate analyses of covariance were used to determine mean between-group differences controlling for baseline level of outcomes. Assessments between the baseline, 8-week, and 16-week follow-up were compared using repeated-measures analysis of variance. A sensitivity analysis (per-protocol analysis) of the primary outcome was conducted with the analysis of covariance including only patients who finished the 16-week follow-up in 2 groups and who followed 80% or more of the prescribed exercise protocol in the intervention group. The chi-square or Fisher exact tests were used to compare frequencies.

Data were documented in case report forms entered into EpiData (version 4.6.0.2, EpiData Association) and analyzed using SPSS (version 24.0, IBM Corp) and GraphPad Prism 8 (GraphPad Software, Inc) software. All statistical tests were 2-sided, and $P<.05$ was considered to be statistically significant.

Results

Participants and Baseline Characteristics

From July to September 2019, we initially screened 77 individuals via social media, of whom 55 patients with AS were randomized in this trial. Among the patients who were randomized, one patient declined to participate due to assignment to the control group. Therefore, a total of 54 patients with AS were assigned as follows: 26 to the intervention group and 28 to the control group. The assignments and patient withdrawal data are presented in Figure 2. Table 1 shows baseline characteristics of the participants.

Figure 2. Treatment assignments and withdrawal in the intention-to-treat population. ITT: intention-to-treat.

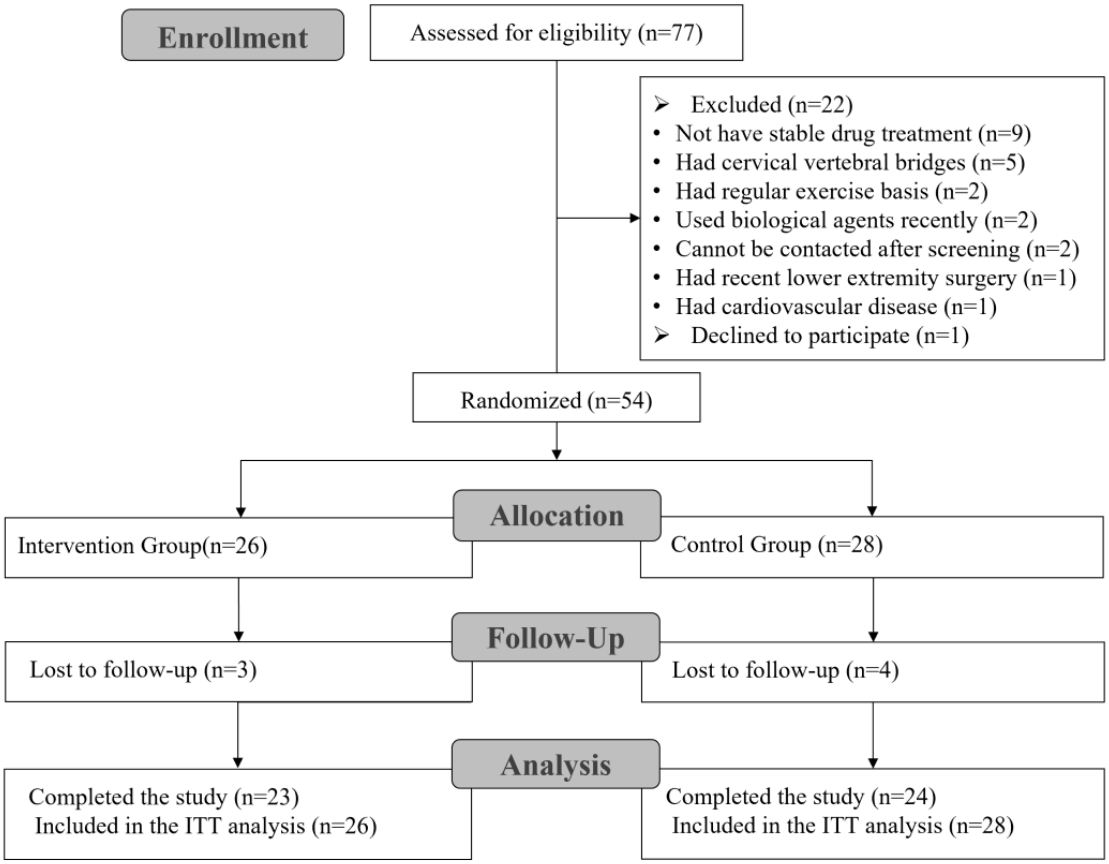


Table 1. Baseline characteristics of patients with ankylosing spondylitis enrolled in the trial.

Characteristics	Intervention group (n=26)	Control group (n=28)
Age at baseline (years), mean (SD)	31.2 (6.3)	33.2 (6.2)
Male gender, n (%)	20 (77)	21 (75)
Married, n (%)	16 (62)	18 (64)
HLA-B27[a] positive, n (%)	24 (92)	25 (89)
Disease duration (years), mean (SD)	10.9 (5.5)	10.1 (6.1)
Height (cm), mean (SD)	173.6 (7.9)	172.0 (8.6)
Weight (kg), mean (SD)	67.7 (11.1)	70.5 (12.1)
BMI (kg/m^2), mean (SD)	22.4 (2.8)	23.8 (3.3)
Past history or current symptoms, n (%)		
AAU[b]	5 (19)	2 (7)
IBD[c]	1 (4)	0
Psoriasis	0	0
Heel pain	6 (23.)	5 (18)
Hip pain	3 (12)	4 (14)
Physical examination, n (%)		
Enthesitis	6 (23)	10 (36)
Peripheral arthritis	4 (15)	3 (11)
PGA[d], mean (SD)	2.9 (1.5)	3.2 (1.1)
PhGA[e], mean (SD)	2.9 (1.2)	3.2 (1.1)
Total pain, mean (SD)	2.1 (1.5)	2.8 (1.4)
Nocturnal pain, mean (SD)	2.5 (2.1)	2.5 (1.5)
ASDAS[f], mean (SD)	1.8 (0.5)	1.7 (0.4)
BASDAI[g], mean (SD)	2.0 (0.8)	2.2 (0.6)
BASMI[h], mean (SD)	1.2 (1.3)	1.4 (1.7)
BASFI[i], mean (SD)	1.1 (1.1)	1.0 (0.8)
ASAS HI[j], mean (SD)	2.7 (2.5)	3.0 (2.9)
ESR[k] (mm/hour)	12.8 (10.0)	10.7 (9.7)
CRP[l] (mg/L)	5.7 (6.5)	3.0 (2.4)
NSAIDs[m], n (%)	18 (69)	22 (79)
csDMARDs[n], n (%)	5 (19)	8 (29)
Corticosteroids, n (%)	0	0

[a]HLA-B27: human leukocyte antigen B27.

[b]AAU: acute anterior uveitis.

[c]IBD: inflammatory bowel disease.

[d]PGA: patient global assessment.

[e]PhGA: physician global assessment.

[f]ASDAS: Ankylosing Spondylitis Disease Activity Score.

[g]BASDAI: Bath Ankylosing Spondylitis Disease Activity Index.

[h]BASMI: Bath Ankylosing Spondylitis Metrology Index.

[i]BASFI: Bath Ankylosing Spondylitis Functional Index.

[j]ASAS HI: Assessment of Spondyloarthritis International Society Health Index.

[k]ESR: erythrocyte sedimentation rate.

[l]CRP: C-reactive protein.

[m]NSAID: nonsteroidal anti-inflammatory drug.

[n]csDMARD: conventional synthetic disease-modifying antirheumatic drug.

Retention and Adherence

The 8-week retention rates were 92% (24/26) for the intervention group and 89% (25/28) for the control group. The 16-week retention rates were 89% (23/26) for the intervention group and 86% (24/28) for the control group. The median adherence rate of the prescribed exercise protocol was 84.2% (IQR 48.7%-97.9%) among all patients assigned to the intervention group. A total of 62% (16/26) of patients in the intervention group followed 80% or more (103/128) of the prescribed exercise protocol. Only 8% (2/26) of patients attended less than 10% (13/128) of the prescribed exercise protocol.

Primary Outcome

The between-group difference for change from baseline to 16-week follow-up of ASDAS was –0.2 (95% CI –0.4 to –0.02, P=.03), indicating more beneficial effect was detected in the intervention group (Multimedia Appendix 3).

Per-protocol analysis excluded 4 participants from the control group and 3 participants from the intervention group who were lost to follow-up and excluded 8 participants who followed less than 80% of the prescribed exercise protocol. The results were similar to primary analyses indicating a significant between-group difference in ASDAS at 16 weeks favoring the intervention (between-group difference for change from baseline, –0.3 [95% CI –0.6 to –0.1], P=.02).

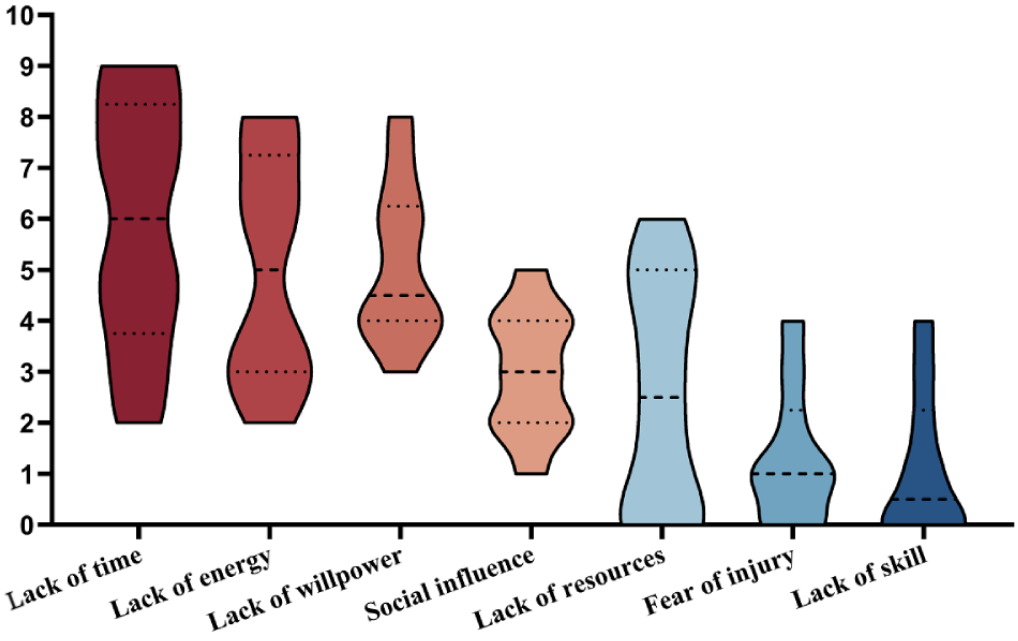
Secondary Outcome

Between-group differences for change from baseline to 16-week follow-up favoring the intervention (Multimedia Appendix 3) were observed in PGA (P<.001), PhGA (P<.001), spinal pain–total pain (P=.004), BASDAI (P=.004), BASDAI–fatigue (P=.04), BASDAI–spinal pain (P=.03), BASDAI–morning stiffness intensity (P=.04), BASDAI–morning stiffness duration (P=.05), BASFI (P=.04), and BASMI (P<.001). Moreover, the frequencies of ASAS HI–motivation (to do anything that requires physical effort) were decreased from 42% (11/26) to 19% (5/26) in the intervention group and from 50% (14/28) to 46% (13/28) in the control group at 16-week follow-up (P=.03), suggesting that this exercise program can increase motivation to do things that require physical effort (Multimedia Appendix 4). For SF-36, significant beneficial effects were detected in physical function, bodily pain, general health, and vitality in the intervention group (Multimedia Appendix 3). In addition, between-group differences for change from baseline of VO_2 max, percentage of body fat, visceral fat area, range of motion of cervical lateral flexion and hip abduction, and back extensor endurance test were significant favoring the intervention group (Multimedia Appendix 5).

Barriers to Being Active

Barriers to being active are presented in Figure 3. The results indicated that lack of time is the most distinct barrier for participants to overcome, followed by lack of energy and lack of willpower.

Figure 3. Barriers to being active. Barriers to being active were analyzed among patients who followed <80% of the prescribed exercise protocol in the intervention group.



Adverse Events

The incidences of adverse events observed in the intervention group and control group were 12% (3/26) and 0, respectively (P=.11), suggesting that no additional significant adverse events were caused by this exercise program. In the intervention group, 1 patient reported ankle pain and 2 patients experienced hip pain during the exercises. About 1 week after the exercise plans were adapted with less jumping movements or other adjustments, the pain disappeared, and these 3 patients all completed the intervention. No serious adverse events occurred during the trial phase in both groups.

Discussion

Principal Findings

In this study, we evaluated the adherence, efficacy, and safety of the wearable technology-assisted combined home-based exercise intervention in AS patients without an exercise habit through a 16-week follow-up. To our knowledge, this is the first randomized controlled trial to implement a wearable technology-assisted home-based exercise intervention in AS.

The wearable technology served as not only a supervisor but also an assessor in this exercise program, as the exercise intensity was determined and the exercise frequency and duration were assessed with the heart rate monitor embedded in the wristband. The dose of exercise reflected by type, intensity, frequency, and duration is decisive in exercise intervention [27]. In previous home-based exercise programs, PA assessments were mostly registered with the use of self-reported questionnaires or diaries, which may be influenced by patient attitudes, poor recall, and giving perceived desired responses rather than accurate ones [28]. Moreover, only class attendance can be registered, and the adherence to exercises within the attended sessions usually remains unclear [8]. In our study, the exercise type was identified, and the exercise intensity, frequency, and duration can be assessed with the heart rate monitor embedded in the wristband. For patient attitudes, a previous perspective survey of patients with axial spondyloarthritis suggested that 82.9% (97/117) of patients agreed with the implementation of exercise with heart rate monitoring, indicating the great need to explore wearable technology in the standard of care [29].

Our study proved that this wearable technology-assisted combined home-based exercise can improve the disease activity of patients with AS. In a study conducted by Hsieh et al [9] comparing the effectiveness of combined home-based exercise and range-of-motion home exercise in patients with AS, no significant improvement was observed in BASDAI in either of the 2 home-based exercise groups, but they posited this may be due to poor adherence (48%). As previous studies reported, adherence for home-based exercise was between 30% and 90%, usually in the lower range; generally, adherence with inpatient exercise was highest, followed by supervised outpatient exercise and home-based exercise [8,9]. Therefore, strategies to improve adherence of home-based exercise should be explored and identified so that the advantages of home-based exercise can be further highlighted. The good adherence rate of the prescribed exercise protocol (84%) and significant improvements of ASDAS and BASDAI in our study indicated that the use of a WAT may improve poor adherence in home-based exercise.

Identifying barriers toward applying exercise is an important issue in the research agenda of the self-management of AS [6,30]. The results of Barriers to Being Active suggested that besides lack of time, lack of energy and lack of willpower were barriers to being active. In exercise, energy and willpower can largely influence patient self-efficacy [31]. Interestingly, in our study, the improvement of BASDAI-fatigue, SF-36–VT, and ASAS HI–motivation indicated that this exercise program can help patients reduce fatigue (energy) and increase motivation to do things that requires physical effort (willpower).

In this wearable technology–assisted home-based exercise, moderate intensity exercise was prescribed considering the paradox of the benefit and harm of activity in AS. On one hand, the PA level of patients with AS should be improved, as lower PA levels are detrimental due to joint instability and loss of strength [11]. On the other hand, enthesitis is a typical feature of AS; a very high PA level is detrimental due to entheseal trauma and moving rapidly from inactivity to activity is dangerous [32]. Although high-intensity exercise is superior to moderate-intensity exercise for improving VO_2 max [33], and previous research indicates that high-intensity exercises can improve the peak oxygen uptake (VO_2 peak) in patients with axial spondyloarthritis over 3 months [34], the high-intensity exercise sessions should be supervised by physiotherapists throughout the trial phase. Without supervisions, it may be dangerous for AS patients considering the increased risk of injuries caused by the precipitate excessive mechanical force applied to joints [32]. Therefore, for AS patients who do not exercise on a regular basis, moderate-intensity exercise may be a more appropriate choice when performing home-based exercise.

Limitations

Some limitations regarding the trial design merit caution. One limitation was that the patients were not blinded to the allocations, which is a common limitation of nonpharmacological treatment. To reduce bias, the primary outcome was evaluated by assessment staff who were unaware of the specific therapeutic regimen. Second, the sample size in this study was not large; however, it was determined that the sample size was sufficient as the effect size in this trial was 0.527, higher than what was set. Finally, to minimize the cofounding effect of drug treatment, only patients treated with nonsteroidal anti-inflammatory drugs and/or disease-modifying antirheumatic drugs were included in this pilot study. We have already initiated a large cohort study to determine the effect and safety of exercise in AS patients treated with different treatment strategies including biological agents.

Conclusion

This pilot study suggests that the wearable technology–assisted combined home-based exercise is feasible and has beneficial effects on disease activity, physical function, spinal mobility, health-related quality of life, range of motion of cervical joints, and back extensor endurance in patients with AS who had no exercise habit.

Acknowledgments

We are extremely grateful to the participants for their support. We appreciate the contribution of the staff of our department for their cooperation throughout the course of this clinical trial.

Authors' Contributions

FH participated in study conception, design, and supervision. YW, XL, and WW were involved in literature search, study design, data collection, statistical analysis, interpretation, and writing and revision of the manuscript. XJ participated in study conception, design, and data interpretation. YS, LH, and LW participated in data collection. YY and SX contributed to the study concept and data interpretation. JZ, JLZ, and WJ were involved in study conception and supervision of the study. All authors revised the manuscript and approved the final report.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study protocol.
[DOCX File , 122 KB - jmir_v24i1e29703_app1.docx]

Multimedia Appendix 2
Ankylosing spondylitis exercise program used in this study.
[DOC File , 36 KB - jmir_v24i1e29703_app2.doc]

Multimedia Appendix 3
Effects of exercise on primary outcome and other disease-related indexes.
[DOCX File , 22 KB - jmir_v24i1e29703_app3.docx]

Multimedia Appendix 4
Assessment of the Spondyloarthritis International Society Health Index was performed to indicate the common difficulties of patients with ankylosing spondylitis.
[DOCX File , 21 KB - jmir_v24i1e29703_app4.docx]

Multimedia Appendix 5
Effects of exercise on VO_2 max, body composition, and range of motion of cervical spine and hip joints.
[DOCX File , 19 KB - jmir_v24i1e29703_app5.docx]

Multimedia Appendix 6
CONSORT-eHEALTH checklist (V 1.6.2).
[PDF File (Adobe PDF File), 103 KB - jmir_v24i1e29703_app6.pdf]

References

1. Sieper J, Poddubnyy D. Axial spondyloarthritis. Lancet 2017 Jul 01;390(10089):73-84. [doi: 10.1016/S0140-6736(16)31591-4] [Medline: 28110981]
2. Millner JR, Barron JS, Beinke KM, Butterworth RH, Chasle BE, Dutton LJ, et al. Exercise for ankylosing spondylitis: an evidence-based consensus statement. Semin Arthritis Rheum 2016 Feb;45(4):411-427 [FREE Full text] [doi: 10.1016/j.semarthrit.2015.08.003] [Medline: 26493464]
3. van der Heijde D, Ramiro S, Landewé R, Baraliakos X, Van den Bosch F, Sepriano A, et al. 2016 update of the ASAS-EULAR management recommendations for axial spondyloarthritis. Ann Rheum Dis 2017 Jun 13;76(6):978-991. [doi: 10.1136/annrheumdis-2016-210770] [Medline: 28087505]
4. Ward MM, Deodhar A, Gensler LS, Dubreuil M, Yu D, Khan MA, et al. 2019 Update of the American College of Rheumatology/Spondylitis Association of America/Spondyloarthritis Research and Treatment Network Recommendations for the Treatment of Ankylosing Spondylitis and Nonradiographic Axial Spondyloarthritis. Arthritis Rheumatol 2019 Oct 22;71(10):1599-1613 [FREE Full text] [doi: 10.1002/art.41042] [Medline: 31436036]

5. Garber CE, Blissmer B, Deschenes MR, Franklin BA, Lamonte MJ, Lee I, et al. Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults. Med Sci Sports Exer 2011;43(7):1334-1359. [doi: 10.1249/MSS.0b013e318213fefb] [Medline: 21694556]
6. Rausch Osthoff A, Niedermann K, Braun J, Adams J, Brodin N, Dagfinrud H, et al. 2018 EULAR recommendations for physical activity in people with inflammatory arthritis and osteoarthritis. Ann Rheum Dis 2018 Sep;77(9):1251-1260. [doi: 10.1136/annrheumdis-2018-213585] [Medline: 29997112]
7. Liang H, Xu L, Tian X, Wang S, Liu X, Dai Y, et al. The comparative efficacy of supervised- versus home-based exercise programs in patients with ankylosing spondylitis: A meta-analysis. Medicine (Baltimore) 2020 Feb;99(8):e19229 [FREE Full text] [doi: 10.1097/MD.0000000000019229] [Medline: 32080122]
8. McDonald MT, Siebert S, Coulter EH, McDonald DA, Paul L. Level of adherence to prescribed exercise in spondyloarthritis and factors affecting this adherence: a systematic review. Rheumatol Int 2019 Feb 16;39(2):187-201 [FREE Full text] [doi: 10.1007/s00296-018-4225-8] [Medline: 30649569]
9. Hsieh L, Chuang C, Tseng C, Wei J, Hsu W, Lin Y. Combined home exercise is more effective than range-of-motion home exercise in patients with ankylosing spondylitis: a randomized controlled trial. Biomed Res Int 2014;2014:398190 [FREE Full text] [doi: 10.1155/2014/398190] [Medline: 25276785]
10. Solomon DH, Rudin RS. Digital health technologies: opportunities and challenges in rheumatology. Nat Rev Rheumatol 2020 Jul 24;16(9):525-535. [doi: 10.1038/s41584-020-0461-x]
11. Jacquemin C, Servy H, Molto A, Sellam J, Foltz V, Gandjbakhch F, et al. Physical activity assessment using an activity tracker in patients with rheumatoid arthritis and axial spondyloarthritis: prospective observational study. JMIR Mhealth Uhealth 2018 Jan 02;6(1):e1 [FREE Full text] [doi: 10.2196/mhealth.7948] [Medline: 29295810]
12. van Riel PL, Zuidema RM, Vogel C, Rongen-van Dartel SA. Patient self-management and tracking: a European experience. Rheum Dis Clin North Am 2019 May;45(2):187-195 [FREE Full text] [doi: 10.1016/j.rdc.2019.01.008] [Medline: 30952392]
13. Choi W, Zheng H, Franklin P, Tulu B. mHealth technologies for osteoarthritis self-management and treatment: a systematic review. Health Informatics J 2019 Sep 27;25(3):984-1003 [FREE Full text] [doi: 10.1177/1460458217735676] [Medline: 29090628]
14. Pellegrini CA, Powell SM, Mook N, DeVivo K, Ehrlich-Jones L. Use of physical activity monitors in rheumatic populations. Curr Rheumatol Rep 2018 Oct 06;20(12):73. [doi: 10.1007/s11926-018-0786-5] [Medline: 30293147]
15. Davergne T, Pallot A, Dechartres A, Fautrel B, Gossec L. Use of wearable activity trackers to improve physical activity behavior in patients with rheumatic and musculoskeletal diseases: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2019 Jun 23;71(6):758-767 [FREE Full text] [doi: 10.1002/acr.23752] [Medline: 30221489]
16. Ji X, Wang Y, Ma Y, Hu Z, Man S, Zhang Y, et al. Improvement of disease management and cost effectiveness in chinese patients with ankylosing spondylitis using a smart-phone management system: a prospective cohort study. Biomed Res Int 2019;2019:2171475 [FREE Full text] [doi: 10.1155/2019/2171475] [Medline: 30931322]
17. Ji X, Hu L, Wang Y, Luo Y, Zhu J, Zhang J, et al. "Mobile health" for the management of spondyloarthritis and its application in China. Curr Rheumatol Rep 2019 Nov 19;21(11):61. [doi: 10.1007/s11926-019-0860-7] [Medline: 31741083]
18. van der Linden S, Valkenburg HA, Cats A. Evaluation of diagnostic criteria for ankylosing spondylitis. a proposal for modification of the New York criteria. Arthritis Rheum 1984 Apr;27(4):361-368. [doi: 10.1002/art.1780270401] [Medline: 6231933]
19. American College of Sports Medicine, American Heart Association. Recommendations for cardiovascular screening, staffing, and emergency policies at health/fitness facilities. Med Sci Sports Exerc 1998 Jun;30(6):1009-1018. [Medline: 9624663]
20. Barriers to Being Active Quiz. Centers for Disease Control and Prevention. URL: https://www.cdc.gov/diabetes/ndep/pdfs/toolkits/road-to-health/road-to-health-barriers-quiz.pdf [accessed 2022-01-03]
21. Lukas C, Landewé R, Sieper J, Dougados M, Davis J, Braun J, Assessment of SpondyloArthritis international Society. Development of an ASAS-endorsed disease activity score (ASDAS) in patients with ankylosing spondylitis. Ann Rheum Dis 2009 Jan 14;68(1):18-24. [doi: 10.1136/ard.2008.094870] [Medline: 18625618]
22. Garrett S, Jenkinson T, Kennedy LG, Whitelock H, Gaisford P, Calin A. A new approach to defining disease status in ankylosing spondylitis: the Bath Ankylosing Spondylitis Disease Activity Index. J Rheumatol 1994 Dec;21(12):2286-2291. [Medline: 7699630]
23. Jenkinson TR, Mallorie PA, Whitelock HC, Kennedy LG, Garrett SL, Calin A. Defining spinal mobility in ankylosing spondylitis (AS). The Bath AS Metrology Index. J Rheumatol 1994 Sep;21(9):1694-1698. [Medline: 7799351]
24. Kiltz U, van der Heijde D, Boonen A, Cieza A, Stucki G, Khan MA, et al. Development of a health index in patients with ankylosing spondylitis (ASAS HI): final result of a global initiative based on the ICF guided by ASAS. Ann Rheum Dis 2015 May 07;74(5):830-835 [FREE Full text] [doi: 10.1136/annrheumdis-2013-203967] [Medline: 24399232]
25. Ware JE, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. conceptual framework and item selection. Med Care 1992 Jun;30(6):473-483. [Medline: 1593914]
26. Gibson A, Wagner D, Heyward V. Advanced Fitness Assessment and Exercise Prescription, 8th Edition. Champaign: Human Kinetics; 2018.

27. Dagfinrud H, Halvorsen S, Vøllestad NK, Niedermann K, Kvien TK, Hagen KB. Exercise programs in trials for patients with ankylosing spondylitis: do they really have the potential for effectiveness? Arthritis Care Res (Hoboken) 2011 Apr 30;63(4):597-603 [FREE Full text] [doi: 10.1002/acr.20415] [Medline: 21452270]
28. Bollen JC, Dean SG, Siegert RJ, Howe TE, Goodwin VA. A systematic review of measures of self-reported adherence to unsupervised home-based rehabilitation exercise programmes, and their psychometric properties. BMJ Open 2014 Jun 27;4(6):e005044-e005044 [FREE Full text] [doi: 10.1136/bmjopen-2014-005044] [Medline: 24972606]
29. Hilberdink B, van der Giesen F, Vliet Vlieland T, van Gaalen F, van Weely S. Supervised group exercise in axial spondyloarthritis: patients' satisfaction and perspective on evidence-based enhancements. Arthritis Care Res (Hoboken) 2020 Jun 25;72(6):829-837 [FREE Full text] [doi: 10.1002/acr.23892] [Medline: 30927518]
30. O'Dwyer T, Monaghan A, Moran J, O'Shea F, Wilson F. Behaviour change intervention increases physical activity, spinal mobility and quality of life in adults with ankylosing spondylitis: a randomised trial. J Physiother 2017 Jan;63(1):30-39 [FREE Full text] [doi: 10.1016/j.jphys.2016.11.009] [Medline: 27989730]
31. Marcus BH, Selby VC, Niaura RS, Rossi JS. Self-efficacy and the stages of exercise behavior change. Res Q Exerc Sport 1992 Mar;63(1):60-66. [doi: 10.1080/02701367.1992.10607557] [Medline: 1574662]
32. Debusschere K, Cambré I, Gracey E, Elewaut D. Born to run: the paradox of biomechanical force in spondyloarthritis from an evolutionary perspective. Best Pract Res Clin Rheumatol 2017 Dec;31(6):887-894. [doi: 10.1016/j.berh.2018.07.011] [Medline: 30509447]
33. Helgerud J, Høydal K, Wang E, Karlsen T, Berg P, Bjerkaas M, et al. Aerobic high-intensity intervals improve VO2max more than moderate training. Med Sci Sports Exerc 2007 Apr;39(4):665-671. [doi: 10.1249/mss.0b013e3180304570] [Medline: 17414804]
34. Sveaas SH, Bilberg A, Berg IJ, Provan SA, Rollefstad S, Semb AG, et al. High intensity exercise for 3 months reduces disease activity in axial spondyloarthritis (axSpA): a multicentre randomised trial of 100 patients. Br J Sports Med 2020 Mar;54(5):292-297. [doi: 10.1136/bjsports-2018-099943] [Medline: 30745314]

Abbreviations

ACSM: American College of Sports Medicine
AS: ankylosing spondylitis
ASAS HI: Spondyloarthritis International Society Health Index
ASDAS: Ankylosing Spondylitis Disease Activity Score
BASDAI: Bath Ankylosing Spondylitis Disease Activity Index
BASFI: Bath Ankylosing Spondylitis Functional Index
BASMI: Bath Ankylosing Spondylitis Metrology Index
DHT: digital health technology
EULAR: European League Against Rheumatism
HRmax: maximal heart rate
PA: physical activity
PGA: patient global assessment
PhGA: physician global assessment
PLA: People’s Liberation Army
SF-36: 36-item Short Form Survey
SpA: spondyloarthritis
SpAMS: Smartphone SpondyloArthritis Management System
VO_2 max: maximal oxygen uptake
WAT: wearable activity tracker

Edited by R Kukafka; submitted 17.04.21; peer-reviewed by P Cheung, F Naeem, S Mukherjee; comments to author 04.10.21; revised version received 23.10.21; accepted 15.11.21; published 18.01.22.

Please cite as:
Wang Y, Liu X, Wang W, Shi Y, Ji X, Hu L, Wang L, Yin Y, Xie S, Zhu J, Zhang J, Jiao W, Huang F
Adherence, Efficacy, and Safety of Wearable Technology–Assisted Combined Home-Based Exercise in Chinese Patients With Ankylosing Spondylitis: Randomized Pilot Controlled Clinical Trial
J Med Internet Res 2022;24(1):e29703
URL: https://www.jmir.org/2022/1/e29703
doi:10.2196/29703
PMID:35040798

Original Paper

Usability, Acceptability, and Satisfaction of a Wearable Activity Tracker in Older Adults: Observational Study in a Real-Life Context in Northern Portugal

Célia Domingos[1,2,3,4], BSc; Patrício Costa[1,2,4], PhD; Nadine Correia Santos[1,2,4,5*], PhD; José Miguel Pêgo[1,2,3,4*], MD, PhD

[1]Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, Braga, Portugal

[2]ICVS/3B's, PT Government Associate Laboratory, Braga/Guimarães, Portugal

[3]iCognitus4ALL – IT Solutions, Braga, Portugal

[4]Clinical Academic Center – 2CA-B, Braga, Portugal

[5]Associação Centro de Medicina P5, School of Medicine, University of Minho, Braga, Portugal

[*]these authors contributed equally

Corresponding Author:
José Miguel Pêgo, MD, PhD
Life and Health Sciences Research Institute (ICVS)
School of Medicine
University of Minho
Largo do Paço
Braga, 4710-057
Portugal
Phone: 351 253 604 800
Email: jmpego@med.uminho.pt

Abstract

Background: The use of activity trackers has significantly increased over the last few years. This technology has the potential to improve the levels of physical activity and health-related behaviors in older adults. However, despite the potential benefits, the rate of adoption remains low among older adults. Therefore, understanding how technology is perceived may potentially offer insight to promote its use.

Objective: This study aimed to (1) assess acceptability, usability, and user satisfaction with the Xiaomi Mi Band 2 in Portuguese community-dwelling older adults in a real-world context; (2) explore the mediating effect of the usability on the relationship between user characteristics and satisfaction; and (3) examine the moderating effect of user characteristics on the relationship between usability and user satisfaction.

Methods: Older adults used the Xiaomi Mi Band 2 over 15 days. The user experience was evaluated through the Technology Acceptance Model 3, System Usability Scale, and User Satisfaction Evaluation Questionnaire. An integrated framework for usability and user satisfaction was used to explore user experience. Statistical data analysis included descriptive data analysis, reliability analysis, confirmatory factor analysis, and mediation and moderation analyses.

Results: A sample of 110 older adults with an average age of 68.41 years (SD 3.11) completed the user experience questionnaires. Mean user acceptance was very high—perceived ease of use: 6.45 (SD 0.78); perceptions of external control: 6.74 (SD 0.55); computer anxiety: 6.85 (SD 0.47); and behavioral intention: 6.60 (SD 0.97). The usability was excellent with an average score of 92.70 (SD 10.73), and user satisfaction was classified as a good experience 23.30 (SD 2.40). The mediation analysis confirmed the direct positive effect of usability on satisfaction (β=.530; P<.01) and the direct negative effect of depression on usability (β=–.369; P<.01). Lastly, the indirect effect of usability on user satisfaction was higher in individuals with lower Geriatric Depression Scale levels.

Conclusions: Findings demonstrate that the Xiaomi Mi Band 2 is suitable for older adults. Furthermore, the results confirmed usability as a determinant of satisfaction with the technology and extended the existing knowledge about wearable activity trackers in older adults.

(J Med Internet Res 2022;24(1):e26652) doi:10.2196/26652

KEYWORDS

user experience; Technology Acceptance Model; health monitoring; fitness trackers; aging; seniors

Introduction

Background

Wearable devices are electronic devices that allow users to automatically track and monitor their physical fitness metrics, including number of steps, level of activity, walking distance, calories burned, heart rate, and sleep patterns [1-4]. Over the last few years, these devices have also become increasingly popular among researchers interested in assessing and intervening on physical activity (PA)–related behaviors in real-world contexts. Wearable devices offer the opportunity to collect objective PA data in a less intrusive and inexpensive manner and provide tailored and personalized interventions in real-time [3,5,6]. In fact, overall, academic and industry research has shown that their use can increase PA levels and promote a healthier lifestyle through real-time self-monitoring of health-related behaviors [3,5,7-10]. However, despite these potential benefits, older adults still show slow technology adoption rates [10,11], possibly because these technologies are mainly developed for a younger target group, without considering health psychology or gerontology theories [7]. Consequently, older adult users may have usability barriers to technology adoption [4,12]. Furthermore, factors associated with normal aging, such as physical and cognitive decline, could limit the ability to use the technology [11].

A better understanding of older adults' intentions to use activity trackers, and examining actual usage behavior, is becoming increasingly relevant; however, only a few studies have been conducted to determine older adults' perceptions [7,10,13,14]. Therefore, this study aimed to understand the user experience and acceptability of an activity tracker (Xiaomi Mi Band 2), throughout daily life activities, in a cohort of community-dwelling older adults.

Theoretical Framework

After carrying out a literature search, 3 major key concepts were identified regarding user experience and technology adoption: technology acceptance, usability, and user satisfaction. Variables regarding user characteristics were also selected, such as cognitive function, mood, and education, which may significantly influence user experience to develop our model. Thus, the theoretical framework was designed to explore older adults' user experience with the Xiaomi Mi Band 2, by combining different theories as next described, while also enabling the examination of the impact of usability and individual characteristics on user satisfaction with the technology.

Technology Acceptance Model

Technology acceptance is an important factor in determining the long-term adoption of activity trackers [3]. The Technology Acceptance Model (TAM) is the most applied theoretical model for evaluating or predicting users' acceptance of new technologies. The TAM was adapted from the Theory of Reasoned Action [15] and was initially developed by Davis [16]. This model assumes that the perceived ease of use (PEOU) and perceived usefulness (PU) are the primary factors influencing an individual's intention to use new technology [3,12,16]. PEOU refers to the degree to which a person perceives how easy it is to use the technology, and PU refers to how using the technology will improve performance [16]. Moreover, PEOU and PU can be influenced by various external factors, including both the device and user characteristics [3,16,17]. The usability seems to be predictive of acceptance regarding the device characteristics because they directly relate to the PEOU and PU and may moderate attitudes and behavioral intentions (BIs) to use a system [3].

The original TAM was extended to TAM 2 by Venkatesh and Davis [18] to explain PU and usage intentions in terms of social influence and cognitive instrumental determinants. Later, Venkatesh and Bala [19] updated the model, including other variables affecting PEOU, such as individual differences (computer self-efficacy, computer anxiety [CANX], and computer playfulness), perceptions of external control (PEC), and system characteristics–related adjustments (perceived enjoyment and objective usability).

System Usability Scale

Initially proposed by John Brooke in 1986, the System Usability Scale (SUS) is the most widely used standardized questionnaire to measure perceived usability [8,17,20,21]. Recent literature shows that several studies extend the TAM by incorporating the SUS [17,22,23]. Although the SUS has been assumed to be unidimensional, recent research reveals that the SUS has 2 subscales—usability and learnability—with items 4 and 10 providing the learnability dimension and the other 8 items the usability dimension [24,25].

According to ISO-9241-11 [26], usability refers to the effectiveness, efficiency, and user satisfaction rating of a product in a specific environment by a particular user for a particular purpose. More precisely, effectiveness refers to which of the system's intended goals can be achieved; efficiency is the effort required for a user to achieve the goals; and satisfaction depends on how comfortable the user feels using the system [8,21,27]. Therefore, usability is a critical factor that directly affects the use and adoption of technology by older adults.

User Satisfaction Evaluation Questionnaire

The literature on technology acceptance has included many model variants and extensions, including user satisfaction as a key indicator of user acceptance [28-34]. Moreover, satisfaction has been described as a predictor of behavior intention [29]. The User Satisfaction Evaluation Questionnaire (USEQ) was initially designed by Gil-Gómez et al [35] to evaluate the satisfaction of the users with virtual rehabilitation systems. Recently, the USEQ was adapted and validated into European Portuguese by Domingos et al [36] to evaluate an activity tracker (Xiaomi Mi Band 2) in older adults, showing psychometric properties consistent with the original version.

User Characteristics

In a theoretical framework developed by Venkatesh and Bala [19], individual differences, such as personality and demographics (eg, traits or individuals' states, gender, and age), were suggested to influence individuals' perceptions of PU and PEOU. Specifically, personality is related to individual differences in cognitive, emotional, and motivational aspects of mental states that result in stable behavioral action [37]. Moreover, personality has been found to affect technology perceptions and acceptance [3,38].

Additionally, older individuals may show age-related declines, including attention, memory, and processing speed, which may further impact how they interact with the technology [3]. The aging process is also associated with a decline in visual faculties, that is, visuospatial functioning, visual acuity, color discrimination, and contrast sensitivity, crucial for learning new information and executing technology-based tasks [39]. Thus, researchers have focused on the impact of cognitive abilities, self-efficacy, and technology-related anxiety in technology acceptance [11]. Lastly, compared with younger adults, the senior population may be more resistant to adopt new technologies due to cultural factors, education, and experience [3].

Research Framework and Hypotheses

This study uses a model based on the SUS to measure usability and the USEQ to measure user satisfaction and incorporate individual characteristics, such as education, mood, and cognitive performance (Figure 1). The design was founded on the basic theory studied to provide a clear causal relationship between the independent variables (exogenous) and the dependent variables (endogenous). The model has 5 variables exploring the user experience with the Xiaomi Mi Band 2 in older adults.

The following hypotheses were formulated:

H1: Usability has a positive effect on satisfaction.

H2: Education has a positive effect on satisfaction.

H3: Education has a positive effect on usability.

H4: Cognition has a positive effect on satisfaction.

H5: Cognition has a positive effect on usability.

H6: Depression has a negative effect on satisfaction.

H7: Depression has a negative effect on usability.

Additionally, user characteristics' potential moderating effect on the direct effect between usability and satisfaction was tested separately for each variable (Figure 2).

Figure 1. Research hypothesis framework.

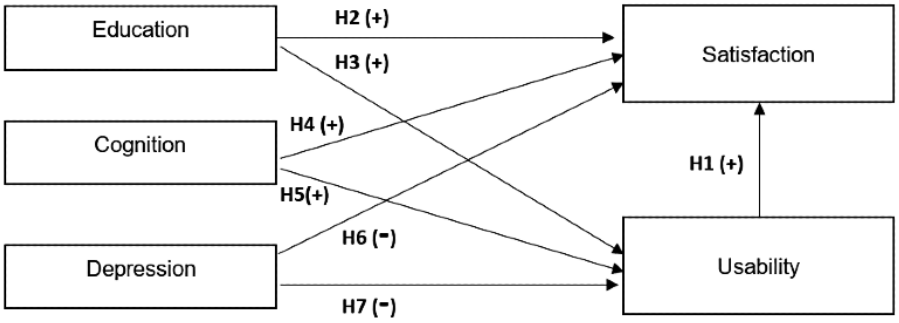
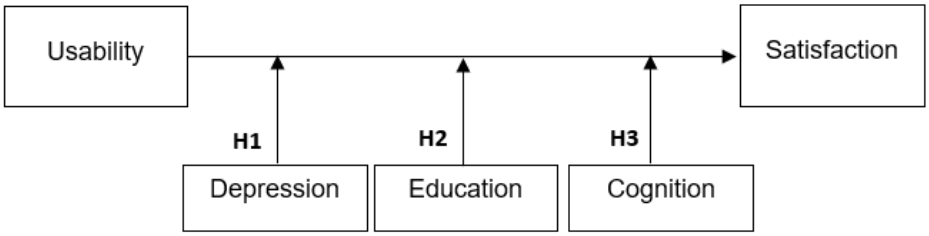


Figure 2. Moderating effect of user characteristics.



Methods

Participants and Research Ethics

A priori sample size calculation and power analysis were performed using G*Power version 3.1.9.3 (Heinrich-Heine-Universität Düsseldorf). Considering that the study is part of a larger project, which used a wearable device to measure and quantify free-living PA in older adults, a total of 120 participants were determined assuming an effect size of 0.32 [40-42], an α of .05, power of 0.95, and dropout rate of 23%. The power analysis for this user experience study was conducted considering the sample size calculated previously and a medium effect size [23] confirmed a power of 0.92. Moreover, the rule of thumb to determine sample size in multiple regression analyses confirmed that the minimum sampling requirements for the analysis were met [43]. Therefore, a total of 120 participants, representative of the general older Portuguese population living in the community within the age group 65-74 years, were recruited from health centers and local gyms in Northern Portugal. The older adults were defined according to the World Health Organization, which considers older people, in developed economies, as those aged 65 or older.

To reduce variability due to the age effect, we used the first 10-year age group in the same way as Eurostat publication—Ageing Europe [44].

The applied exclusion criteria comprised inability to understand informed consent; diagnosed neuropsychiatric and neurodegenerative disorders; or disability that limited independent walking, visual, auditory, or fine motor skills. Participants having previous experience with other wearable activity trackers were not excluded from the study. A final sample of 110 participants was enrolled in the study. The study ran from April 2018 to July 2019.

The study was conducted according to the Helsinki Declaration and approved by the local and ethics committees (Approval Number 42-2018), developed in compliance with the new General Data Protection Regulation, and approved by the Portuguese Data Protection Authority (Approval Number 11286/2017). Study goals and assessments were explained during screening procedures. All participants provided written informed consent before study enrollment, which included consent to their data processing.

Data Collection and Instruments

A baseline characterization was performed through a sociodemographic questionnaire, and a neuropsychological evaluation to obtain mood (Geriatric Depression Scale [GDS]) [45] and global cognitive profiles (Mini-Mental State Examination [MMSE]) [46]. For screening "cognitive impairment" via the MMSE, the following cutoff values were used: individuals with no education, <15 points; 1-11 years of school completed, <22 points; and >11 years of school completed, <27 points [47]. For assessment of the presence of depressive symptomatology via the GDS, the cutoff value considered was a total number of depressive symptoms over 11 [48].

To assess the users' experience, the Xiaomi Mi Band 2 was provided to participants that should be worn continuously for over 15 days, while performing their normal daily activities. The wearable was returned after the testing period for data analysis. In other studies, testing periods range from 3 to 7 days [13,49-51]; because a 7-day testing period corresponds to a short-term user experience, we decided to extend this period to 15 days. Upon completing the usage period, participants were also asked to provide information about their user experience. The TAM 3 [19] was used to collect information about technology acceptance, the SUS [25] for perceived usability, and the USEQ [35,36] for user satisfaction.

Xiaomi Mi Band 2

The selection of wearable activity tracker was based on a review of several different commercially available devices on the market [8,52,53]. The selection criteria included their popularity in the health tracking device market, availability, continuous monitoring of PA without a smartphone, price, battery life, various data captured via sensors, and ability to export data. The Xiaomi Mi Band 2 was selected because, at the study time, it offered the best price-quality ratio, had an estimated battery life of almost 30 days, was ergonomic, accessible, easy to operate, and did not require continuous communication with a smartphone. The system combines sensors that allow the objective assessment of daily free-living PA, with its algorithms calculating steps, intensity, energy expenditure, and distance traveled [49,53,54].

Technology Acceptance Model

The TAM 3 was adapted to the context of the use of activity tracking technologies by older adults, and the key dimensions of acceptance were investigated using the following constructs: PEOU, PEC, CANX, BI, and USE. PEOU was measured using all 4 items adapted from the TAM 3; PEC using 2; CANX using 3, BI and USE were measured using the only item on the original scale. Multimedia Appendix 1 presents a list of items for all the constructs. TAM items were measured on a 7-point Likert scale, starting from "1=strongly disagree" to "7=strongly agree". The mean scores of each item were computed and the mean of means of each construct was calculated and used to perform statistical analysis [19].

System Usability Scale

The SUS is a 10-item questionnaire, consisting of 5 positive and 5 negative statements, with the 5 responses for each statement ranging from "5=strongly agree" to "1=strongly disagree" (Multimedia Appendix 2). The SUS score is calculated by taking 1 from all the scores on odd-numbered items and subtracting 5 from the even-numbered items scores. The sum of the scores is then multiplied by 2.5 to give an overall SUS score, and range from 0 (extremely poor usability) to 100 (excellent usability) [21,25]. The value of 68 is considered the average for the SUS score; a score above or less than 68 is considered above average or below average, respectively [55,56]. The grade rankings of scores proposed by Bangor et al [56] were here used to provide a more meaningful basis for the SUS score interpretation.

User Satisfaction Evaluation Questionnaire

The USEQ is a 6-item questionnaire with a 5-point Likert Scale (Multimedia Appendix 3). The total score ranges from 6 (poor satisfaction) to 30 (excellent satisfaction). All items are affirmative, except item 5, which is a negative item. The numerical value of the affirmative items is used to calculate the score. The negative item subtracts the numerical value of the response from 6 and then adds this result to the total score. The USEQ score is evaluated using the following classification: poor (0-5), fair (5-10), good (10-15), very good (15-20), or excellent (20-25) satisfaction [35,36].

Statistical Analysis

Overview

The statistical analysis was organized to address the following aims: (1) explore the mediating effect of the usability on the relationship between user characteristics and satisfaction; and (2) examine the moderating effect of user characteristics on the relationship between usability and user satisfaction. Briefly, the statistical analysis was performed according to the following steps: (1) descriptive statistics; (2) instruments' psychometric proprieties; (3) structural equation modeling (SEM); and (4) moderation analysis.

Descriptive Statistics

Descriptive data analysis was performed using IBM SPSS Statistics (version 26) to depict the characteristics of the study. Descriptive statistics, including frequency, percentage, mean, standard deviation, minimum, maximum, skewness, and kurtosis, were calculated for each variable. Normality was considered adequate if absolute values for skewness and kurtosis were above 3.0 and 10.0, respectively [57,58]. The percentage of missing values across the variables was analyzed. Methods for handling missing data were not applied because there were no missing data.

Instruments' Validation

Before structural modeling, the measurement model of latent variables for their dimensionality/structure and reliability was assessed.

Confirmatory factor analysis (CFA) was conducted using JASP (version 0.11.1; JASP Team, University of Amsterdam) to examine the structure of the SUS (used to measure usability) and USEQ (used to measure user satisfaction). Variables with factor loadings above 0.4 were included. To assess the goodness of fit of the model, the following indices and thresholds were applied: chi-square (χ^2, P>.05), χ^2/degrees of freedom (*df*) ratio (≤3), Comparative Fit Index (CFI ≥0.90), Tucker–Lewis Index (TLI ≥0.90), Goodness-of-Fit Index (GFI ≥0.90), root mean squared error of approximation (RMSEA <0.08), and standardized root mean squared residual (SRMR ≤0.08) [59-61].

Reliability analysis was performed using IBM SPSS Statistics (version 26) to analyze the internal consistency of item responses of the SUS and USEQ instruments. Reliability was estimated using the McDonald omega (ω_t) coefficient [62,63]. Given ordinal response format items, the McDonald omega coefficient (ω_t) provides more accurate estimates of reliability than Cronbach α [62,64,65]. Coefficients values over 0.70 are considered indicators of satisfactory item homogeneity [65,66].

Structural Equation Modeling

SEM was applied to check the hypothesis relationship between the proposed factors that directly and indirectly influence older adult's user satisfaction (structural model) with technology. SEM allows to analyze the structural relationship between measured variables and latent variables. The derived scores for usability and user satisfaction were supported by CFA.

Data were analyzed using IBM SPSS AMOS (version 25) and the parameters were estimated by the maximum likelihood method. The significance level of 5% was used as a threshold for the research proposition testing. To determine whether the model was reasonable and acceptable, the following indices were considered: χ^2, χ^2/df ratio, CFI, TLI, GFI, and RMSEA. The criteria for an acceptable model fit were the same as those reported for the CFA.

To assess multicollinearity, the inspection of the correlation matrix of the predictor variables (education, MMSE, GDS, and usability) and the analysis of the variance inflation factor (VIF) and tolerance were performed (IBM SPSS Statistics, version 26). The tolerance values close to 1 were considered as an indicator of low multicollinearity, whereas a value close to 0 as a potential indicator of collinearity problem [67,68]. Moreover, VIF=1 was considered an indicator that the independent variables are not correlated, and 1<VIF<5 an indicator that the variables are moderately correlated with each other [67].

Moderation Analysis

Moderation analysis was performed to examine whether the relationship between usability (predictor) and user satisfaction (outcome variable) depended on user characteristics (moderator). The analysis was performed using the MedMod package in jamovi (version 1.2.27; The jamovi Project) software. The significance of the interaction term of usability on user satisfaction at specific values (–1 SD, mean, +1 SD) of GDS, education, and MMSE (moderators) was assessed, exploring when the effect of usability on user satisfaction depends on the level of the moderating test variable.

Results

Study Participants

A total of 110 participants completed the final assessment after the testing period. Table 1 summarizes the demographic, mood, and global cognitive characteristics of the sample. Participants had a mean age of 68.41 (SD 3.11) years, and 45.5% (50/110) were identified as males. The mean years of formal education were 7.95 (SD 5.38).

Table 1. Characteristics of the study participants (N=110).

Characteristics	Values
Gender	
Male, n (%)	50 (45.5)
Age (years), mean (SD)	68.41 (3.11)
64-70, n (%)	73 (66.4)
≥70, n (%)	37 (33.6)
Education (years), mean (SD)	7.95 (5.38)
1-4, n (%)	58 (52.7)
5-11, n (%)	24 (21.8)
≥12, n (%)	28 (25.5)
MMSE[a] (total score), mean (SD)	26.95 (2.00)
22-27, n (%)	41 (37.3)
≥27, n (%)	69 (62.7)
GDS[b] (total score), mean (SD)	6.05 (4.58)
>11, n (%)	17 (15.5)

[a]MMSE: Mini-Mental State Examination.

[b]GDS: Geriatric Depression Scale.

Instruments' Descriptive Statistics

The results of descriptive statistics for the instruments (TAM 3, SUS, and USEQ) are presented in Tables 2-4, respectively. The skewness and kurtosis values indicate some degree of non-normality. In reality, most behavioral research data do not follow univariate normal distributions [69,70]. Moreover, the results reveal a severe violation of normality for the following items and constructs: USEQ 1, SUS 1, SUS 3, SUS 5, SUS 9, PEOU 3, PEC, and CANX. Thus, these were excluded from further path analysis.

Table 2. Descriptive statistics for Technology Acceptance Model 3 items.

Items	Range	Mean (SD)	Skewness	Kurtosis
Perceived Ease of Use (PEOU)				
PEOU 1	2-7	6.28 (1.08)	–1.62	2.25
PEOU 2	1-7	6.06 (2.01)	–1.97	2.22
PEOU 3	3-7	6.84 (0.60)	–4.37	20.92
PEOU 4	3-7	6.60 (0.92)	–2.42	5.11
PEOU score	3.50-7.00	6.45 (0.78)	–1.48	1.54
PEOU final	3.67-7.00	6.31 (0.94)	–1.25	0.24
Perceptions of External Control (PEC)				
PEC 1	3-7	6.55 (0.97)	–2.38	5.19
PEC 2	4-7	6.94 (0.41)	–6.84	46.91
PEC score	4.00-7.00	6.74 (0.55)	–2.62	7.34
Computer Anxiety (CANX)				
CANX 1	6-7	6.99 (0.10)	–10.49	110.00
CANX 2	1-7	6.86 (0.83)	–6.72	45.64
CANX 3	2-7	6.71 (0.97)	–3.49	11.44
CANX score	4.33-7.00	6.85 (0.47)	–3.55	12.72
Behavioral intention (BI)				
BI	1-7	6.60 (0.97)	–3.00	10.84
USE (hours)	13-24	23.85 (1.12)	–8.99	85.16

Table 3. Descriptive statistics for System Usability Scale items.

Items	Range	Mean (SD)	Skewness	Kurtosis
1	1-5	4.71 (0.65)	–2.82	9.87
2	1-5	1.44 (1.03)	2.40	4.70
3	2-5	4.92 (0.36)	–5.81	40.38
4	1-5	1.38 (1.04)	2.51	4.75
5	1-5	4.85 (0.56)	–4.47	23.01
6	1-5	1.28 (0.83)	3.12	9.26
7	1-5	3.58 (0.78)	–2.15	4.88
8	1-5	1.30 (0.92)	3.06	8.16
9	3-5	4.86 (0.46)	–3.42	10.71
10	1-5	1.41 (1.08)	2.52	4.95
System Usability Scale score	55-100	92.70 (10.73)	–1.61	1.77

Table 4. Descriptive statistics for User Satisfaction Evaluation Questionnaire items.

Items	Range	Mean (SD)	Skewness	Kurtosis
2	3-5	4.82 (0.47)	–2.66	6.46
3	2-5	4.65 (0.71)	–2.21	4.50
4	2-5	4.47 (0.75)	–1.30	0.98
5	1-5	4.65 (0.93)	–2.71	6.15
6	1-5	1.28 (0.83)	3.12	9.26
User Satisfaction Evaluation Questionnaire score	14-25	23.30 (2.40)	–1.81	2.99

Instruments' Psychometric Proprieties

As reported by Domingos et al [36], the CFA supported the conceptual unidimensionality of the USEQ (χ^2_4=1.83, *P*=.12, χ^2/df=1.83; CFI=0.973, TLI=0.931, GFI=0.977, RMSEA=0.087, SRMR=0.038). Furthermore, the CFA for the SUS showed satisfactory values for the following indexes: CFI=0.816, GFI=0.928, and SRMR=0.074. The fit indices for the model are presented in Table 5.

Regarding internal consistency, for the SUS questionnaire reliability was calculated only with items included in path analysis (SUS 2, SUS 4, SUS 6, SUS 7, SUS 8, SUS 10). Moreover, the USEQ showed acceptable reliability (Cronbach α=.677; McDonald ω=0.722), as reported by Domingos et al [36]. The McDonald ω coefficients showed acceptable values for the SUS and USEQ questionnaires ranging from 0.712 to 0.722, respectively.

Table 5. Confirmatory factor analysis for instruments.

Fit indices	User Satisfaction Evaluation Questionnaire	System Usability Scale
χ^2	7.313	30.074
df	4	9
χ^2/df	1.83	3.34
P value	.120	<.001
Comparative Fit Index	0.973	0.816
Tucker–Lewis Index	0.931	0.694
Goodness-of-Fit Index	0.977	0.928
Root mean squared error of approximation	0.087	0.146
Standardized root mean squared residual	0.038	0.074

Users' Experience

The high ratings of the TAM 3 indicate excellent technology acceptance by the participants. Overall, the average ratings for user experience with the Xiaomi Mi Band 2 were 6.45 (SD 0.78) for PEOU, 6.74 (SD 0.55) for PEC, 6.85 (SD 0.47) for CANX, and 6.60 (SD 0.97) for BI. Furthermore, the participants reported an average of 23.85 (SD 1.12) hours of use per day (Table 2). These results indicate that participants found that the Xiaomi Mi Band 2 is an easy-to-use and easy-to-control device, potentially perceiving its usefulness regarding health benefits and having the intention to use it in the future.

Regarding usability, the overall SUS score ranged from 55 to 100 (mean [SD] 92.70 [10.73]), with 96% (106/110) of the participants reporting a score above the acceptability baseline of the SUS. Moreover, 45.5% (50/110) of the participants classified the activity tracker achieved as best imaginable (Table 6). Thus, these results suggest that the Xiaomi Mi Band 2 is a usable wearable activity tracker among older adults.

Finally, all participants reported a user satisfaction experience above the USEQ baseline value defined as a good experience, with a mean USEQ score of 23.30 (SD 2.40; Table 4). Moreover, 85.5% (94/110) of the participants rated the satisfaction with the Xiaomi Mi Band 2 as excellent (Table 6). Still, despite older adults reporting good satisfaction with the device, concerns were noted regarding the clarity of the technology's information.

Table 6. User experience classification for usability and satisfaction (N=110).

Classification	Value, n (%)
Usability (System Usability Scale)	
Ok	8 (7.3)
Good	16 (14.5)
Excellent	36 (32.7)
Best imaginable	50 (45.5)
Satisfaction (User Satisfaction Evaluation Questionnaire)	
Good	2 (1.8)
Very good	14 (12.7)
Excellent	94 (85.5)

The Structural Equation Modeling for User Satisfaction

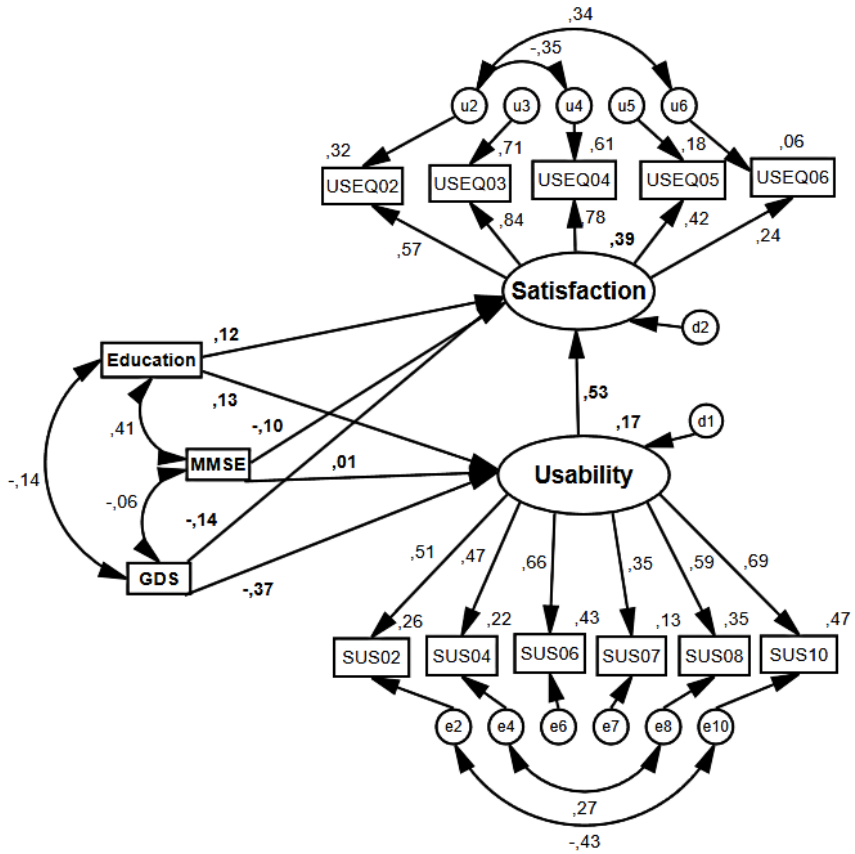
Table 7 shows fit indexes for the structural model, showing acceptable values for the χ^2/df (1.67) and RMSEA (0.079) indexes and values slightly less than the threshold for a good model fit for the following indexes: GFI=0.880, TLI=0.818, and CFI=0.868. Based on these indexes, the model has a moderate acceptable fit.

The path diagram of the model is presented in Figure 3. Coefficients within paths are standardized coefficients from regressions. Table 8 summarizes the results of hypothesis testing, including standardized coefficients and significance levels. Specifically, results show that usability was significantly and positively associated with user satisfaction (β=.530; P<.01), thereby supporting Hypotheses 1. By contrast, depression was significantly and negatively associated with usability (β=–.369; P<.01), supporting Hypotheses 7.

Table 7. Fit indices for the hypothesized model.

Model fit index	Value
χ^2	110.475
df	66
χ^2/df	1.67
P value	<.001
Goodness-of-Fit Index	0.880
Tucker–Lewis Index	0.818
Comparative Fit Index	0.868
Root mean squared error of approximation	0.079

Figure 3. Path diagram for the research model. GDS: Geriatric Depression Scale; MMSE: Mini-Mental State Examination; SUS: System Usability Scale; USEQ: User Satisfaction Evaluation Questionnaire.



Individual characteristics (education, cognition, and depression) collectively explained 16.8% of usability variance. Furthermore, individual characteristics and usability collectively explained 39.1% of the variance in satisfaction. Specifically, depression negatively impacted usability and satisfaction, with a significant effect on usability (β=–.369; *P*<.01); while, regarding education, a positive, but not significant, usability and satisfaction effect was observed (education > satisfaction: β=–.121; *P*<.23; education > usability: β=–.130; *P*<.25). Despite confirming the theoretical model, most research hypotheses were not statistically proven with adequate goodness of fit. Nonetheless, usability seems to be a strong predictor of user satisfaction.

Considering the possible multicollinearity issues in the SEM, the absolute values of correlation coefficients were calculated and ranged from 0.009 to 0.45. The tolerance values ranged from 0.82 to 0.87 and the VIF values from 1.15 to 1.22, indicating that no independent variable is in a perfect linear function with other any independent variable.

Table 8. Results of hypothesis testing based on standardized path coefficients for the research model.

Hypothesis	Estimate	Standard error	Critical ratio	*P* value
H1: Usability > Satisfaction	0.530	0.089	3.008	.003
H2: Education > Satisfaction	0.121	0.005	1.194	.23
H3: Education > Usability	0.130	0.011	1.147	.25
H4: Cognition > Satisfaction	–0.098	0.013	–0.999	.32
H5: Cognition > Usability	0.011	0.029	0.104	.92
H6: Depression > Satisfaction	–0.140	0.006	–1.376	.17
H7: Depression > Usability	–0.369	0.014	–3.010	.003

Moderation Analysis

Moderating Effect of the GDS

Usability significantly predicted satisfaction (β=.43; *P*<.001; Table 9). The interaction effect of usability × GDS was not significant (β=–.028; *P*<.06); however, because the *P*-value is approximately .05, we can conclude there is a tendency to infer that the effect of the satisfaction is dependent on GDS levels. The simple slopes of the interaction at –1 SD, mean, and +1 SD of GDS are plotted in Figure 4. Results indicate a significant association for high and low values of low GDS, respectively, in the same direction (β=.30, *P*<.001; β=.56, *P*<.001; Table 10).

Moreover, the effect of usability on user satisfaction through the GDS was higher in individuals with lower GDS levels.

Table 9. Estimates for the moderating effect of the GDS[a] and usability in the prediction of user satisfaction.

Variable	Estimate	Standard error	Z	*P* value
Usability	0.43	0.082	5.28	<.001
GDS	–0.012	0.009	–1.34	.18
H1: Usability × GDS	–0.028	0.015	–1.90	.06

[a]GDS: Geriatric Depression Scale.

Figure 4. Simple slope plot for the moderating effect of Geriatric Depression Scale (GDS) and usability in the prediction of user satisfaction.

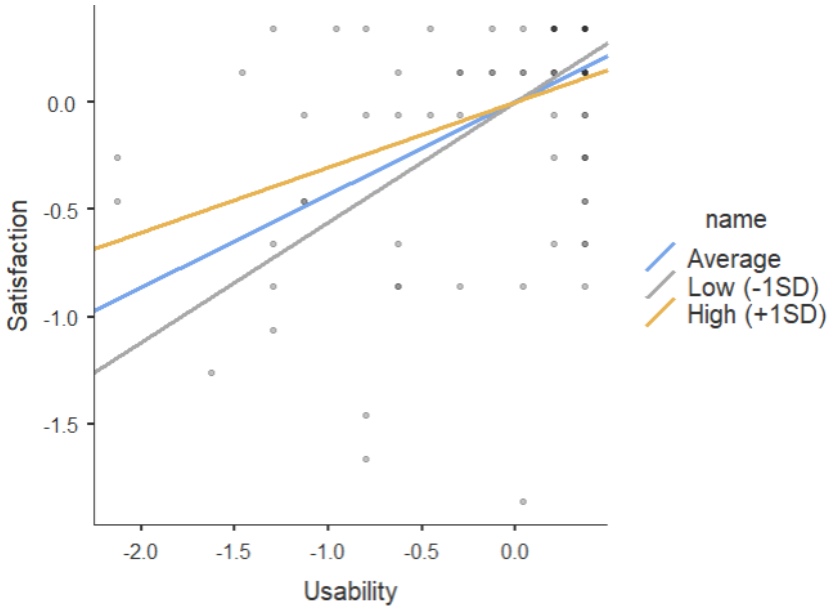


Table 10. Effect of the usability on satisfaction at different levels of the GDS[a].

Effect	Estimate	Standard error	Z	*P* value
Average	0.43	0.083	5.22	<.001
Low (–1 SD)	0.56	0.135	4.16	<.001
High (+1 SD)	0.30	0.070	4.36	<.001

[a]GDS: Geriatric Depression Scale.

Moderating Effect of Education

Usability significantly predicted satisfaction (β=.36; *P*<.001; Table 11). However, the interaction effect of usability × education in the direct path between usability and user satisfaction was not significant (β=3.63 × 10^{-4}; *P*<.98). Results from simple slope estimates for the effect of usability on satisfaction indicated that education did not moderate the relationship between these variables (Table 12). Moreover, the interaction plot (Figure 5) showed no difference in simple slopes at –1 SD, mean, and +1 SD.

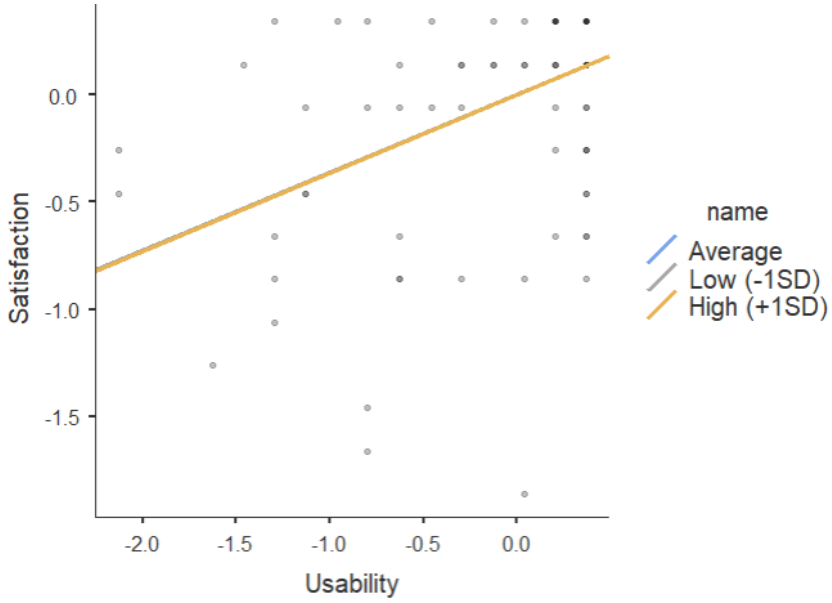
Table 11. Estimates for the moderating effect of education and usability in the prediction of user satisfaction.

Variable	Estimate	Standard error	Z	*P* value
Usability	0.36	0.070	5.23	<.001
Education	0.004	0.008	0.522	.60
H2: Usability × Education	3.63 × 10^{-4}	0.017	0.022	.98

Table 12. Effect of the usability on satisfaction at different levels of education.

Effect	Estimate	Standard error	Z	*P* value
Average	0.364	0.067	5.23	<.001
Low (–1 SD)	0.362	0.098	3.71	<.001
High (+1 SD)	0.366	0.128	2.85	.004

Figure 5. Simple slope plot for the moderating effect of education and usability in the prediction of user satisfaction.



Moderating Effect of the MMSE

The interaction effect of usability × MMSE in the direct path between usability and user satisfaction was not significant (β=.014; *P*<.66; Table 13). The simple slopes of the interaction at –1 SD, mean, and +1 SD of the GDS are plotted in Figure 6.

Results indicate a positive relationship between usability and satisfaction for both low (β=.352; *P*<.001) and high (β=.407; *P*<.001) MMSE levels (Table 14). Additionally, the results suggested that the indirect effect of usability on user satisfaction through the MMSE is higher for individuals with higher levels of MMSE.

Table 13. Estimates for the moderating effect of the MMSE[a] and usability in the prediction of user satisfaction.

Variable	Estimate	Standard error	Z	*P* value
Usability	0.380	0.070	5.46	<.001
MMSE	–0.008	0.020	–0.387	.70
H3: Usability × MMSE	0.014	0.031	0.445	.66

[a]MMSE: Mini-Mental State Examination.

Figure 6. Simple slope plot for moderating effect of Mini-Mental State Examination (MMSE) and usability in the prediction of user satisfaction.

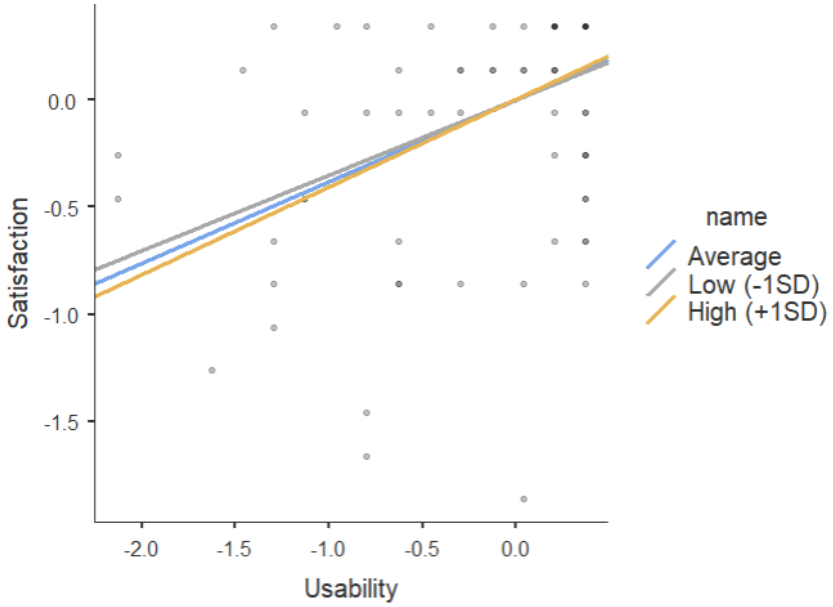


Table 14. Effect of the usability on satisfaction at different levels of Mini-Mental State Examination (MMSE).

Effect	Estimate	Standard error	Z	*P* value
Average	0.380	0.070	5.45	<.001
Low (–1 SD)	0.352	0.079	4.44	<.001
High (+1 SD)	0.407	0.105	3.87	<.001

Discussion

Principal Findings

In recent years, wearable activity trackers are part of a rapidly growing trend in biomedical research and medicine [13,71,72]. These devices have been employed in behavior change interventions due to their potential to motivate individuals to comply with a daily activity goal [13,71]. Because most older adults have insu cient levels of PA, these technologies may be especially beneficial in middle-aged and older age groups. Nonetheless, it is reported that only 16% of activity tracker owners are 55-64 years of age and 7% over 65 [4], indicating that possibly these devices may not be feasible or acceptable to older adults [73]. Therefore, it is necessary to understand how older adults perceive these new technologies. Thus, to better understand potential barriers to using these technologies, the acceptability, usability, and user satisfaction experience with the Xiaomi Mi Band 2 were examined in a population of older adults.

Results from users' experience indicate an excellent technology acceptance with high ratings of the TAM 3 in all constructs, including PEOU, PEC, CANX, and BI. Previously, Puri et al [49] found a moderate level of acceptance (65%) for the Xiaomi Mi Band 2 among the Canadian community-dwelling older adults, where, interestingly, participants reported a significantly higher acceptance rate for the Xiaomi Mi Band 2 when compared with Microsoft Band. In our study, no other devices were tested, and thus, did not allow for any comparison between wearable devices.

Concerning usability, all participants scored their experience above the acceptability baseline for the SUS. Thus, these results indicate that the Xiaomi Mi Band 2 has excellent usability for older adults in this specific context. Participants also reported a user satisfaction experience above the USEQ baseline value defined as a good experience, suggesting excellent user satisfaction. Nonetheless, such a large score on the SUS, mean 92.70 (SD 10.73), was surprising. In the study by Liang et al [8], the Xiaomi Mi Band 2 was one of the devices that achieved the highest score among several selected wearable devices with distinct market performance, but its mean SUS score was 65.12 (SD 14.73). Possible explanations range from the intrinsic motivation to use the device and how the device is supplied; therefore, such aspects should be evaluated in future studies.

This study also examined factors influencing user satisfaction with the Xiaomi Mi Band 2, based on the proposed theoretical framework. The hypothetical model was supported by moderate acceptable fit indices values (χ^2/df=1.67, GFI=0.880, TLI=0.818, CFI=0.868, and RMSEA=0.079). Furthermore, 2 of the testing hypotheses were proven. Overall, results indicate that usability is a significant predictor of user satisfaction (β=.530; *P*<.01), which, in turn, was negatively affected by depression symptoms (β=–.369; *P*<.01). The model shows that individual characteristics explain 16.8% of the usability variance and 39.1% of the variance in satisfaction collectively with usability. Specifically, a significant negative effect of depression on usability was found (β=–.369; *P*<.01).

Additionally, user characteristics' potential moderating effect on the interaction between usability and user satisfaction was

examined. Results suggested that the GDS moderates the usability effect on user satisfaction, and the effect is higher in individuals with lower GDS levels. However, we did not observe significant moderating effects for education ($\beta=3.63 \times 10^{-4}$; $P<.98$) and MMSE ($\beta=.014$; $P<.66$) on the interaction between usability and user satisfaction, contrary to our expectations. Future research should explore additional moderating effects through the user characteristics, including personal traits as well as motivational and cultural aspects to enable a better understanding of the factors that may influence user satisfaction and consequently facilitate technology adoption.

Overall, our results align with a recent study investigating the impact of depressive symptoms on web user experience measures, indicating that mood may be a factor influencing technology usability [74]. Additionally, recent research investigating the relationship between user perceptions and user characteristics has shown that older adults demonstrate positive attitudes toward mobile technologies and report technologies' complexity. User characteristics, such as age, processing speed, and attention, significantly influence older adults' usage behavior. Furthermore, the education level was found to be positively correlated with the diversity of use. Probably, individuals with higher education levels are typically more motivated to accept new concepts. The authors also mentioned that the usability problems could be attributed to poor memory, decreased vision, and poor literacy, thus older adults tended to perceive the technologies as difficult to use [39].

Beyond the proposed research framework of the study, we aimed to use an integrated TAM and user satisfaction, similar to other studies [29,31,34]. However, due to the severe violation of normality observed in TAM 3 constructs, we cannot integrate the TAM in path analysis. Nonetheless, previous research has shown a significant influence of PEOU on user satisfaction, with the latter proposed to be a key predictor of BI [32-34]. Additionally, Chao [29] showed that perceived enjoyment, effort expectancy, and performance expectancy have a significantly positive effect on satisfaction; thus, it would have been relevant to include these variables to predict satisfaction.

Regarding usability, Venkatesh et al [19,75] theorized that PEOU is affected by the objective usability of a specific system only after a direct experience with the system, where perceptions about the PEOU are determined solely by usability features, which in turn form the basis for acceptance or rejection. Moreover, if the system has higher objective usability, it means that system that is easy to use. Several studies suggested that usability is a determinant of PEOU [19,23,75].

Regarding study limitations, our sample is not representative of the entire older population because we used a convenience sample. Therefore, findings cannot be widely generalizable. Moreover, the population sample is more homogenous than the wider population on the common factors, possibly leading to attenuation in correlations or erroneous correlations among variables [76,77]. Although we have a minimum sample size adequate for the estimation method (>100 participants), the SEM is a large-sample technique [78]. Therefore, future studies should have a larger and more heterogeneous sample to obtain sufficiently accurate estimates, although our study had a larger sample size compared with previous ones [13,49-51]. A further limitation is that the user experience was assessed for a specific wearable activity tracker (Xiaomi Mi Band 2), and therefore, is not representative of the full range of devices currently available on the market. Moreover, the testing period was limited to 15 days. Short-term technology acceptance may not be indicative of long-term acceptance, as research indicates that use of activity trackers tend to drop after the first few weeks [1,49], with short timeframes also making it difficult to determine the impact of the novelty effect (defined as a person's subjective "first responses to a technology, not the patterns of usage that will persist over time as the product ceases to be new" [79]). Moreover, research suggests that the declining novelty effect could be a reason for many activity tracker users discontinuing their use. Recently, Shin et al [80] explored the effect of novelty in the early stages (<3 months) of activity tracker adoption, as well as the motivation factors for sustained activity tracker use in the long term (>6 months). Findings reveal that the use beyond the novelty period is determined by intrinsic and extrinsic motivations. Finally, we selected the SUS for the usability evaluation because it is the most widely used questionnaire to measure perceived usability; however, this instrument does not comprise all of the concepts regarding usability. For instance, there are several different standards (eg, ISO-9241-11 [26], ISO/IEC 9126 [81]) and conceptual models to evaluate usability. Shackel [82] reported on the 4 important characteristics of usability, namely, effectiveness, learnability, flexibility, and attitude, and the Nielsen model (1993) [83] gave 5 subattributes of usability, namely, learnability, efficiency, memorability, errors, and satisfaction [84,85]. Therefore, there is a need for future studies evaluating key dimensions of usability.

Conclusions

In conclusion, while there is a pressing need for studies to include other devices currently on the market and evaluate longer-term use, our study extended on the existing research providing valuable insight into the use of wearable activity trackers among older adults. First, a significant contribution of this work was to demonstrate the relevance of usability as an important factor influencing user satisfaction, which probably has an impact on technology acceptance and on the intention to use activity trackers. However, we were not able to predict BI in our structural model. Furthermore, our results emphasize the need to consider strategies to minimize the usability barriers to technology adoption in older adults. In addition, system designers should provide systems that address these concerns, and the researchers must ensure that selected systems adequately address the usability issues to be effectively implemented in clinical and research settings. Second, our study investigated the impact of user characteristics as moderating factors influencing the relationship between usability and user satisfaction and found that depression symptoms have a significant influence on older adults' perception of using technology. However, other individual differences/personal user characteristics should be examined, and the identified moderating effects should be taken into consideration when implementing strategies trying to promote technology adoption. Finally, our results suggested that the Xiaomi Mi Band 2 is a

suitable wearable activity tracker for older adults to use in real-life context.

Acknowledgments

Financial support for this work was provided by FEDER funds through the Operational Programme Competitiveness Factors – COMPETE and National Funds through FCT under the project POCI-01-0145-FEDER-007038 (UIDB/50026/2020, and UIDP/50026/2020), by the projects NORTE-01-0145-FEDER-000013 and NORTE-01-0145-FEDER-000023 (supported by the North Portugal Regional Operational Programme [NORTE 2020], under the Portugal 2020 [P2020] Partnership Agreement, through the European Regional Development Fund [FEDER]), by POCI-01-0145-FEDER-016428 (supported by the Operational Programme Competitiveness and Internationalization [COMPETE 2020] and the Regional Operational Program of Lisbon and National Funding through Portuguese Foundation for Science and Technology [FCT, Portugal]), and by the Portuguese North Regional Operational Programme (ON.2 – O Novo Norte, under the National Strategic Reference Framework [QREN], through FEDER). The work was also developed under the scope of the 2CA-Braga Grant of the 2017 Clinical Research Projects. CD was supported by a combined PhD scholarship from FCT and the company iCognitus4ALL - IT Solutions, Lda, Braga, Portugal (Grant number PD/BDE/127831/2016).

Authors' Contributions

CD was responsible for conceptualization, data curation, formal analysis, investigation, methodology, and writing (original draft, review, and editing). PSC was responsible for formal analysis, methodology, and writing (review and editing). NCS and JMP were responsible for funding acquisition, supervision, and writing (review and editing). All authors reviewed and approved the final version of the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Measurement items of TAM 3. TAM: Technology Acceptance Model.
[DOCX File , 14 KB - jmir_v24i1e26652_app1.docx]

Multimedia Appendix 2
Measurement items of USEQ. USEQ: User Satisfaction Evaluation Questionnaire.
[DOCX File , 13 KB - jmir_v24i1e26652_app2.docx]

Multimedia Appendix 3
Measurement items of SUS. SUS: System Usability Scale.
[DOCX File , 13 KB - jmir_v24i1e26652_app3.docx]

References

1. Shin G, Jarrahi MH, Fei Y, Karami A, Gafinowitz N, Byun A, et al. Wearable activity trackers, accuracy, adoption, acceptance and health impact: A systematic literature review. J Biomed Inform 2019 May;93:103153 [FREE Full text] [doi: 10.1016/j.jbi.2019.103153] [Medline: 30910623]
2. Shih P, Han K, Poole E, Rosson M, Carroll J. Use and adoption challenges of wearable activity trackers. In: Proceedings of the iConference. 2015 Presented at: iConference; March 15, 2015; Newport Beach, CA p. 1-12 URL: https://www.researchgate.net/publication/268746784_Use_and_adoption_challenges_of_wearable_activity_trackers
3. Rupp MA, Michaelis JR, McConnell DS, Smither JA. The role of individual differences on perceptions of wearable fitness device trust, usability, and motivational impact. Appl Ergon 2018 Jul;70:77-87. [doi: 10.1016/j.apergo.2018.02.005] [Medline: 29866329]
4. Steinert A, Haesner M, Steinhagen-Thiessen E. Activity-tracking devices for older adults: comparison and preferences. Univ Access Inf Soc 2017 Apr 8;17(2):411-419. [doi: 10.1007/s10209-017-0539-7]
5. Turner-McGrievy G, Jake-Schoffman DE, Singletary C, Wright M, Crimarco A, Wirth MD, et al. Using Commercial Physical Activity Trackers for Health Promotion Research: Four Case Studies. Health Promot Pract 2019 May;20(3):381-389 [FREE Full text] [doi: 10.1177/1524839918769559] [Medline: 29618233]
6. Alley S, van Uffelen JG, Schoeppe S, Parkinson L, Hunt S, Power D, et al. Efficacy of a computer-tailored web-based physical activity intervention using Fitbits for older adults: a randomised controlled trial protocol. BMJ Open 2019 Dec 23;9(12):e033305 [FREE Full text] [doi: 10.1136/bmjopen-2019-033305] [Medline: 31874890]

7. Seifert A, Schlomann A, Rietz C, Schelling HR. The use of mobile devices for physical activity tracking in older adults' everyday life. Digit Health 2017;3:2055207617740088 [FREE Full text] [doi: 10.1177/2055207617740088] [Medline: 29942617]
8. Liang J, Xian D, Liu X, Fu J, Zhang X, Tang B, et al. Usability Study of Mainstream Wearable Fitness Devices: Feature Analysis and System Usability Scale Evaluation. JMIR Mhealth Uhealth 2018 Nov 08;6(11):e11066 [FREE Full text] [doi: 10.2196/11066] [Medline: 30409767]
9. Maher C, Ryan J, Ambrosi C, Edney S. Users' experiences of wearable activity trackers: a cross-sectional study. BMC Public Health 2017 Nov 15;17(1):880 [FREE Full text] [doi: 10.1186/s12889-017-4888-1] [Medline: 29141607]
10. Kononova A, Li L, Kamp K, Bowen M, Rikard R, Cotten S, et al. The Use of Wearable Activity Trackers Among Older Adults: Focus Group Study of Tracker Perceptions, Motivators, and Barriers in the Maintenance Stage of Behavior Change. JMIR Mhealth Uhealth 2019 Apr 05;7(4):e9832 [FREE Full text] [doi: 10.2196/mhealth.9832] [Medline: 30950807]
11. Berkowsky R, Sharit J, Czaja SJ. Factors Predicting Decisions About Technology Adoption Among Older Adults. Innov Aging 2018 Jan;2(1):igy002 [FREE Full text] [doi: 10.1093/geroni/igy002] [Medline: 30480129]
12. Preusse KC, Mitzner TL, Fausset CB, Rogers WA. Older Adults' Acceptance of Activity Trackers. J Appl Gerontol 2017 Feb;36(2):127-155 [FREE Full text] [doi: 10.1177/0733464815624151] [Medline: 26753803]
13. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of Commercially Available Wearable Activity Trackers Among Adults Aged Over 50 and With Chronic Illness: A Mixed-Methods Evaluation. JMIR Mhealth Uhealth 2016 Jan 27;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
14. Schlomann A, Seifert A, Rietz C. Relevance of Activity Tracking With Mobile Devices in the Relationship Between Physical Activity Levels and Satisfaction With Physical Fitness in Older Adults: Representative Survey. JMIR Aging 2019 Mar 06;2(1):e12303 [FREE Full text] [doi: 10.2196/12303] [Medline: 31518263]
15. LaCaille L. Theory of Reasoned Action. In: Gellman MD, Turner JR, editors. Encyclopedia of Behavioral Medicine. New York, NY: Springer; 2013:1964-1967.
16. Davis FD. Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. MIS Quarterly 1989 Sep;13(3):319-340. [doi: 10.2307/249008]
17. Revythi A, Tselios N. Extension of technology acceptance model by using system usability scale to assess behavioral intention to use e-learning. Educ Inf Technol 2019 Feb 1;24(4):2341-2355. [doi: 10.1007/s10639-019-09869-4]
18. Venkatesh V, Davis FD. A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. Management Science 2000 Feb;46(2):186-204. [doi: 10.1287/mnsc.46.2.186.11926]
19. Venkatesh V, Bala H. Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences 2008 May;39(2):273-315. [doi: 10.1111/j.1540-5915.2008.00192.x]
20. Lewis JR. The System Usability Scale: Past, Present, and Future. International Journal of Human–Computer Interaction 2018 Mar 30;34(7):577-590. [doi: 10.1080/10447318.2018.1455307]
21. Brooke J. SUS - A quick and dirty usability scale. Usability Evaluation in Industry. 1996. URL: https://www.researchgate.net/publication/228593520_SUS_A_quick_and_dirty_usability_scale [accessed 2022-01-11]
22. Pande T, Saravu K, Temesgen Z, Seyoum A, Rai S, Rao R, et al. Evaluating clinicians' user experience and acceptability of , a smartphone application for tuberculosis in India. Mhealth 2017 Jul 27;3:30-30 [FREE Full text] [doi: 10.21037/mhealth.2017.07.01] [Medline: 28828377]
23. M Scholtz B, Mahmud I, T. R. Does Usability Matter? An Analysis of the Impact of Usability on Technology Acceptance in ERP Settings. IJIKM 2016;11:309-330. [doi: 10.28945/3591]
24. Lewis JR, Sauro J. The factor structure of the System Usability Scale. In: Human Centered Design. Berlin, Heidelberg: Springer; 2009:94-103.
25. Martins AI, Rosa AF, Queirós A, Silva A, Rocha NP. European Portuguese Validation of the System Usability Scale (SUS). Procedia Computer Science 2015;67:293-300. [doi: 10.1016/j.procs.2015.09.273]
26. International Organization for Standardization. Ergonomics of Human-System Interaction — Part 11: Usability: Definitions and Concepts. 2018. URL: https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:en [accessed 2022-01-11]
27. Bevan N, Carter J, Earthy J, Geis T, Harker S. New ISO standards for usability, usability reports and usability measures. In: Human-Computer Interaction. Theory, Design, Development and Practice. Cham, Switzerland: Springer; 2016:268-278.
28. Dalcher I, Shine J. Extending the New Technology Acceptance Model to Measure the End User Information Systems Satisfaction in a Mandatory Environment: A Bank's Treasury. Technology Analysis & Strategic Management 2003 Dec;15(4):441-455. [doi: 10.1080/0953732030001360 33]
29. Chao C. Factors Determining the Behavioral Intention to Use Mobile Learning: An Application and Extension of the UTAUT Model. Front Psychol 2019 Jul 16;10:1652 [FREE Full text] [doi: 10.3389/fpsyg.2019.01652] [Medline: 31379679]
30. Wixom BH, Todd PA. A Theoretical Integration of User Satisfaction and Technology Acceptance. Information Systems Research 2005 Mar;16(1):85-102. [doi: 10.1287/isre.1050.0042]
31. Mather D, Caputi P, Jayasuriya R. Is the technology acceptance model a valid model of user satisfaction of information technology in environments where usage is mandatory? ACIS 2002 Proceedings. 2002. URL: https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1149&context=acis2002 [accessed 2022-01-11]

32. Ghobakhloo M, Zulkifli B, Aziz F. The interactive model of user information technology acceptance and satisfaction in small and medium-sized enterprises. European Journal of Economics, Finance and Administrative Sciences 2010;19(1):7-27 [FREE Full text]
33. Ho K, Ho C, Chung M. Theoretical integration of user satisfaction and technology acceptance of the nursing process information system. PLoS One 2019 Jun 4;14(6):e0217622 [FREE Full text] [doi: 10.1371/journal.pone.0217622] [Medline: 31163076]
34. He L, Kim H, Gong H. The Influence of Consumer and Product Characteristics on Intention to Repurchase of Smart band. International Journal of Asia Digital Art and Design Association 2017;21(1):13-18. [doi: 10.20668/adada.21.1_13]
35. Gil-Gómez JA, Manzano-Hernández P, Albiol-Pérez S, Aula-Valero C, Gil-Gómez H, Lozano-Quilis J. USEQ: A Short Questionnaire for Satisfaction Evaluation of Virtual Rehabilitation Systems. Sensors (Basel) 2017 Jul 07;17(7):1589 [FREE Full text] [doi: 10.3390/s17071589] [Medline: 28686174]
36. Domingos C, Costa PS, Santos NS, Pêgo JM. European Portuguese Version of the User Satisfaction Evaluation Questionnaire (USEQ): Transcultural Adaptation and Validation Study. JMIR Mhealth Uhealth 2021 Jun 29;9(6):e19245 [FREE Full text] [doi: 10.2196/19245] [Medline: 34185018]
37. Montag C, Panksepp J. Primary Emotional Systems and Personality: An Evolutionary Perspective. Front Psychol 2017;8:464 [FREE Full text] [doi: 10.3389/fpsyg.2017.00464] [Medline: 28443039]
38. Svendsen GB, Johnsen JK, Almås-Sørensen L, Vittersø J. Personality and technology acceptance: the influence of personality factors on the core constructs of the Technology Acceptance Model. Behaviour & Information Technology 2013 Apr;32(4):323-334. [doi: 10.1080/0144929x.2011.553740]
39. Li Q, Luximon Y. Understanding Older Adults Post-Adoption Usage Behavior and Perceptions of Mobile Technology. International Journal of Design 2018;12(3):93-110 [FREE Full text]
40. Hogan CL, Mata J, Carstensen LL. Exercise holds immediate benefits for affect and cognition in younger and older adults. Psychol Aging 2013 Jun;28(2):587-594 [FREE Full text] [doi: 10.1037/a0032634] [Medline: 23795769]
41. Gajewski PD, Falkenstein M. Physical activity and neurocognitive functioning in aging - a condensed updated review. Eur Rev Aging Phys Act 2016 Jan 21;13(1):1 [FREE Full text] [doi: 10.1186/s11556-016-0161-3] [Medline: 26865880]
42. Lautenschlager NT, Cox K, Cyarto EV. The influence of exercise on brain aging and dementia. Biochim Biophys Acta 2012 Mar;1822(3):474-481 [FREE Full text] [doi: 10.1016/j.bbadis.2011.07.010] [Medline: 21810472]
43. Green SB. How Many Subjects Does It Take To Do A Regression Analysis. Multivariate Behav Res 1991 Jul 01;26(3):499-510. [doi: 10.1207/s15327906mbr2603_7] [Medline: 26776715]
44. Eurostat. Ageing Europe. Looking at the Lives of Older People in the EU. 2020. URL: https://ec.europa.eu/eurostat/documents/3217494/11478057/KS-02-20-655-EN-N.pdf/9b09606c-d4e8-4c33-63d2-3b20d5c19c91?t=1604055531000 [accessed 2022-01-11]
45. Yesavage JA, Brink T, Rose TL, Lum O, Huang V, Adey M, et al. Development and validation of a geriatric depression screening scale: A preliminary report. Journal of Psychiatric Research 1982 Jan;17(1):37-49. [doi: 10.1016/0022-3956(82)90033-4]
46. Guerreiro M, Silva AA, Botelho MF, Leitão O, Castro-Caldas A, Garcia C. Adaptação à população portuguesa da tradução do Mini Mental State Examination (MMSE). Revista Portuguesa de Neurologia 1994;1(9):9-10.
47. Santana I, Duro D, Lemos R, Costa V, Pereira M, Simões MR, et al. [Mini-Mental State Examination: Screening and Diagnosis of Cognitive Decline, Using New Normative Data]. Acta Med Port 2016 Apr;29(4):240-248 [FREE Full text] [doi: 10.20344/amp.6889] [Medline: 27349775]
48. Pocinho M, Farate C, Dias C, Lee T, Yesavage J. Clinical and Psychometric Validation of the Geriatric Depression Scale (GDS) for Portuguese Elders. Clinical Gerontologist 2009 Feb 23;32(2):223-236. [doi: 10.1080/07317110802678680]
49. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User Acceptance of Wrist-Worn Activity Trackers Among Community-Dwelling Older Adults: Mixed Method Study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173 [FREE Full text] [doi: 10.2196/mhealth.8211] [Medline: 29141837]
50. Vooijs M, Alpay LL, Snoeck-Stroband JB, Beerthuizen T, Siemonsma PC, Abbink JJ, et al. Validity and usability of low-cost accelerometers for internet-based self-monitoring of physical activity in patients with chronic obstructive pulmonary disease. Interact J Med Res 2014 Oct 27;3(4):e14 [FREE Full text] [doi: 10.2196/ijmr.3056] [Medline: 25347989]
51. Farina N, Lowry RG. Older adults' satisfaction of wearing consumer-level activity monitors. J Rehabil Assist Technol Eng 2017 Oct 31;4:2055668317733258 [FREE Full text] [doi: 10.1177/2055668317733258] [Medline: 31186939]
52. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the Validity of Current Mainstream Wearable Devices in Fitness Tracking Under Various Physical Activities: Comparative Study. JMIR Mhealth Uhealth 2018 Apr 12;6(4):e94 [FREE Full text] [doi: 10.2196/mhealth.9754] [Medline: 29650506]
53. El-Amrawy F, Nounou MI. Are Currently Available Wearable Devices for Activity Tracking and Heart Rate Monitoring Accurate, Precise, and Medically Beneficial? Healthc Inform Res 2015 Oct;21(4):315-320 [FREE Full text] [doi: 10.4258/hir.2015.21.4.315] [Medline: 26618039]
54. Mičková E, Machová K, Daďová K, Svobodová I. Does Dog Ownership Affect Physical Activity, Sleep, and Self-Reported Health in Older Adults? Int J Environ Res Public Health 2019 Sep 11;16(18):3355 [FREE Full text] [doi: 10.3390/ijerph16183355] [Medline: 31514379]

55. Brooke J. SUS: a retrospective. Journal of Usability Studies 2013;8(2):29-40 [FREE Full text]
56. Bangor A, Kortum P, Miller J. Determining what individual SUS scores mean: adding an adjective rating scale. The Journal of Usability Studies 2009;4(3):114-123 [FREE Full text]
57. Kline RB. Principles and Practice of Structural Equation Modeling. New York, NY: Guilford Publications; 2015.
58. Diop EB, Zhao S, Duy TV. An extension of the technology acceptance model for understanding travelers' adoption of variable message signs. PLoS One 2019 Apr 25;14(4):e0216007 [FREE Full text] [doi: 10.1371/journal.pone.0216007] [Medline: 31022277]
59. Hooper D, Coughlan J, Mullen M. Structural equation modelling: guidelines for determining model fit. Electronic Journal on Business Research Methods 2008;6(1):53-60 [FREE Full text]
60. Schermelleh-Engel K, Moosbrugger H, Müller H. Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit measures. Methods of Psychological Research 2003;8(2):23-74 [FREE Full text]
61. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal 1999 Jan;6(1):1-55. [doi: 10.1080/10705519909540118]
62. Şimşek GG, Noyan F. McDonald's ωt, Cronbach's α, and Generalized θ for Composite Reliability of Common Factors Structures. Communications in Statistics - Simulation and Computation 2013 Oct;42(9):2008-2025. [doi: 10.1080/03610918.2012.689062]
63. Trizano-Hermosilla I, Alvarado JM. Best Alternatives to Cronbach's Alpha Reliability in Realistic Conditions: Congeneric and Asymmetrical Measurements. Front Psychol 2016 May 26;7:769 [FREE Full text] [doi: 10.3389/fpsyg.2016.00769] [Medline: 27303333]
64. Costello A, Osborne J. Best practices in exploratory factor analysis: four recommendations for getting the most from your analysis. Practical Assessment 2005;10(7):1-9 [FREE Full text] [doi: 10.4135/9781412995627.d8]
65. Gadermann AM, Guhn M, Zumbo BD. Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. Practical Assessment, Research, and Evaluation 2012;17(3):1-13 [FREE Full text]
66. Cronbach LJ. Coefficient alpha and the internal structure of tests. Psychometrika 1951 Sep;16(3):297-334. [doi: 10.1007/bf02310555]
67. Shrestha N. Detecting Multicollinearity in Regression Analysis. AJAMS 2020 Jan 15;8(2):39-42. [doi: 10.12691/ajams-8-2-1]
68. Senaviratna NAMR, A. Cooray TMJ. Diagnosing Multicollinearity of Logistic Regression Model. AJPAS 2019 Oct 01:1-9. [doi: 10.9734/ajpas/2019/v5i230132]
69. Curran PJ, West SG, Finch JF. The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis. Psychological Methods 1996 Mar;1(1):16-29. [doi: 10.1037/1082-989x.1.1.16]
70. Norman G. Likert scales, levels of measurement and the "laws" of statistics. Adv Health Sci Educ Theory Pract 2010 Dec 10;15(5):625-632. [doi: 10.1007/s10459-010-9222-y] [Medline: 20146096]
71. Bassett DR, Freedson PS, John D. Wearable Activity Trackers in Clinical Research and Practice. Kinesiology Review 2019 Feb;8(1):11-15. [doi: 10.1123/kr.2018-0060]
72. Keogh A, Dorn JF, Walsh L, Calvo F, Caulfield B. Comparing the Usability and Acceptability of Wearable Sensors Among Older Irish Adults in a Real-World Context: Observational Study. JMIR Mhealth Uhealth 2020 Apr 20;8(4):e15704 [FREE Full text] [doi: 10.2196/15704] [Medline: 32310149]
73. Lyons EJ, Swartz MC, Lewis ZH, Martinez E, Jennings K. Feasibility and Acceptability of a Wearable Technology Physical Activity Intervention With Telephone Counseling for Mid-Aged and Older Adults: A Randomized Controlled Pilot Trial. JMIR Mhealth Uhealth 2017 Mar 06;5(3):e28 [FREE Full text] [doi: 10.2196/mhealth.6967] [Medline: 28264796]
74. Thielsch M, Thielsch C. Depressive symptoms and web user experience. PeerJ 2018;6:e4439 [FREE Full text] [doi: 10.7717/peerj.4439] [Medline: 29507832]
75. Venkatesh V. Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model. Information Systems Research 2000 Dec;11(4):342-365. [doi: 10.1287/isre.11.4.342.11872]
76. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychological Methods 1999 Sep;4(3):272-299. [doi: 10.1037/1082-989x.4.3.272]
77. Wolf EJ, Harrington KM, Clark SL, Miller MW. Sample Size Requirements for Structural Equation Models: An Evaluation of Power, Bias, and Solution Propriety. Educ Psychol Meas 2013 Dec 09;76(6):913-934 [FREE Full text] [doi: 10.1177/0013164413495237] [Medline: 25705052]
78. Kyriazos TA. Applied Psychometrics: Sample Size and Sample Power Considerations in Factor Analysis (EFA, CFA) and SEM in General. PSYCH 2018;09(08):2207-2230. [doi: 10.4236/psych.2018.98126]
79. Sung J, Christensen HI, Grinter RE. Robots in the wild: Understanding long-term use. In: HRI '09: Proceedings of the 4th ACM/IEEE international conference on Human robot interaction. 2009 Presented at: HRI09: International Conference on Human Robot Interaction; March 9-13, 2009; La Jolla, CA p. 45-52. [doi: 10.1145/1514095.1514106]
80. Shin G, Feng Y, Jarrahi MH, Gafinowitz N. Beyond novelty effect: a mixed-methods exploration into the motivation for long-term activity tracker use. JAMIA Open 2019 Apr;2(1):62-72 [FREE Full text] [doi: 10.1093/jamiaopen/ooy048] [Medline: 31984346]

81. International Organization for Standardization. Software Engineering — Product Quality — Part 1: Quality Model. 2001. URL: https://www.iso.org/standard/22749.html [accessed 2022-01-11]
82. Shackel B. Usability—context, framework, definition, design, and evaluation. In: Shackel B, Richardson S, editors. Human Factors for Informatics Usability. Cambridge, MA: Cambridge University Press; 1991:21-38.
83. Nielsen J. Usability Engineering. San Diego, CA: Academic Press; 1993.
84. Weichbroth P. Usability attributes revisited: a time-framed knowledge map. In: Proceedings of the Federated Conference on Computer Science and Information Systems. New York, NY: IEEE; 2018 Presented at: Federated Conference on Computer Science and Information Systems; September 9-12, 2018; Poznań, Poland p. 1005-1008 URL: https://annals-csis.org/Volume_15/drp/pdf/137.pdf [doi: 10.15439/2018F137]
85. Madan A, Dubey SK. Usability evaluation methods: a literature review. International Journal of Engineering Science and Technology 2012 Feb;4(2):590-599 [FREE Full text]

Abbreviations

BI: behavioral intention
CANX: computer anxiety
CFA: confirmatory factor analysis
CFI: Comparative Fit Index
GDS: Geriatric Depression Scale
GFI: Goodness of Fit Index
MMSE: Mini-Mental State Examination
PA: physical activity
PEC: perceptions of external control
PEOU: perceived ease of use
PU: perceived usefulness
RMSEA: Root Mean Squared Error of Approximation
SEM: structural equation modeling
SRMR: Standardized Root Mean Squared Residual
SUS: System Usability Scale
TAM: Technology Acceptance Model
TLI: Tucker–Lewis Index
USEQ: User Satisfaction Evaluation Questionnaire
VIF: variance inflation factor

Edited by R Kukafka; submitted 19.12.20; peer-reviewed by N Georgi, S Mukherjee, T Russell-Rose; comments to author 15.02.21; revised version received 24.03.21; accepted 05.07.21; published 26.01.22.

Please cite as:
Domingos C, Costa P, Santos NC, Pêgo JM
Usability, Acceptability, and Satisfaction of a Wearable Activity Tracker in Older Adults: Observational Study in a Real-Life Context in Northern Portugal
J Med Internet Res 2022;24(1):e26652
URL: https://www.jmir.org/2022/1/e26652
doi:10.2196/26652
PMID:35080503

Original Paper

Design of an Integrated Acceptance Framework for Older Users and eHealth: Influential Factor Analysis

Jiyeon Yu[1*], MDes; Angelica de Antonio[1*], MSc, PhD; Elena Villalba-Mora[2,3*], MSc, PhD

[1]Research Group on Human-Computer Interaction and Advanced Interactive Systems, Madrid HCI Lab, Universidad Politécnica de Madrid, Madrid, Spain

[2]Centre for Biomedical Technology (CTB), Universidad Politécnica de Madrid, Madrid, Spain

[3]Centro de Investigación Biomédica en Red en Bioingeniería, Biomateriales y Nanomedicina (CIBER-BBN), Madrid, Spain

[*]all authors contributed equally

Corresponding Author:
Elena Villalba-Mora, MSc, PhD
Centre for Biomedical Technology (CTB)
Universidad Politécnica de Madrid
Pozuelo de Alarcón
Madrid, 28223
Spain
Phone: 34 910679250
Email: elena.villalba@upm.es

Abstract

Background: eHealth and telehealth play a crucial role in assisting older adults who visit hospitals frequently or who live in nursing homes and can benefit from staying at home while being cared for. Adapting to new technologies can be difficult for older people. Thus, to better apply these technologies to older adults' lives, many studies have analyzed the acceptance factors for this particular population. However, there is not yet a consensual framework that can be used in further development and to search for solutions.

Objective: This paper aims to present an integrated acceptance framework (IAF) for older users' acceptance of eHealth based on 43 studies selected through a systematic review.

Methods: We conducted a 4-step study. First, through a systematic review in the field of eHealth from 2010 to 2020, the acceptance factors and basic data for analysis were extracted. Second, we conducted a thematic analysis to group the factors into themes to propose an integrated framework for acceptance. Third, we defined a metric to evaluate the impact of the factors addressed in the studies. Finally, the differences among the important IAF factors were analyzed according to the participants' health conditions, verification time, and year.

Results: Through a systematic review, 731 studies were found in 5 major databases, resulting in 43 (5.9%) selected studies using the PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) methodology. First, the research methods and acceptance factors for eHealth were compared and analyzed, extracting a total of 105 acceptance factors, which were grouped later, resulting in an IAF. A total of 5 dimensions (ie, personal, user–technology relational, technological, service-related, and environmental) emerged, with a total of 23 factors. In addition, we assessed the quality of evidence and then conducted a stratification analysis to reveal the more appropriate factors depending on the health condition and assessment time. Finally, we assessed the factors and dimensions that have recently become more important.

Conclusions: The result of this investigation is a framework for conducting research on eHealth acceptance. To elaborately analyze the impact of the factors of the proposed framework, the criteria for evaluating the evidence from the studies that have the extracted factors are presented. Through this process, the impact of each factor in the IAF has been presented, in addition to the framework proposal. Moreover, a meta-analysis of the current status of research is presented, highlighting the areas where specific measures are needed to facilitate eHealth acceptance.

(J Med Internet Res 2022;24(1):e31920) doi:10.2196/31920

KEYWORDS

eHealth; older people; older user; health technology; acceptance factors; adoption; acceptance framework; systematic review; thematic analysis; influential factor analysis; mobile phone

Introduction

Background

The world's population is aging, and this phenomenon will affect the health care system for older people in the future, and we need to be prepared [1]. During the COVID-19 pandemic, it has been revealed that older people are an especially risky group, and public health authorities have advised them to stay safely at home [2]. This makes it harder for many older people to visit hospitals or health care facilities, and the need for eHealth services to provide health care at home has been increasing. The medical services offered by an internet-based platform have the advantage of increased equality in access to medical services during any type of crisis.

eHealth has been a World Health Organization priority since 2005. It defined eHealth as "a health-related field including medical and health services, health surveillance, health literature, health education, knowledge, and research" and has provided international reports on eHealth readiness [3]. In recent years, eHealth has been increasingly used as a generic term that covers a variety of mobile health (mHealth), telemedicine, and telehealth services, as well as eHealth data management [4]. eHealth is becoming an important solution for people who need to consistently manage their health even at home and receive immediate professional medical services by providing low-cost and high-quality health care [3].

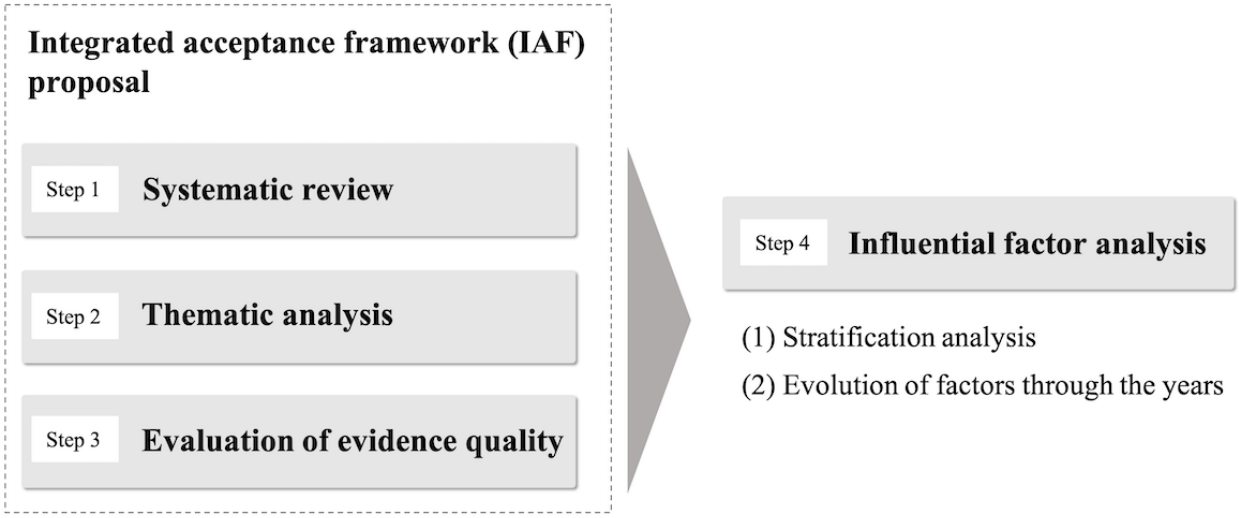
The development of the Internet of Things (IoT) has contributed to advances in the eHealth field. IoT is a technology that allows physical objects, devices, and computers to interact using networks to collect and exchange data [5]. Ambient assisted living in health care facilities using IoT is designed to help older adults' lives. This allows older adults with chronic diseases to measure blood pressure, glucose, electrocardiogram, and body temperature, which need to be monitored every day, and the data can be immediately shared with medical facilities [6]. Recently, advances in sensors and machine learning have made it possible to better perceive and understand the daily lives of older adults. This could lead to the development of eHealth technologies that monitor daily health conditions, share information with health care facilities, and respond to emergencies. The improvement of these technologies can relieve the social burden of aging and accelerate the transition to personalized digital health care that can meet the needs of individuals seeking independent living [7].

However, older people's acceptance, adoption, and use of technology have lagged behind that of younger people. In addition, older people may have low eHealth literacy or low ability to access, evaluate, and use health information to make medical decisions [8]. Nonetheless, as their health concerns and health care needs are higher than for any other age group, some studies have confirmed that a growing number of people from this demographic segment are accepting the technology and are willing to use it in the future [9]. The first step toward applying eHealth technology in the future and bringing it into real life is to identify the factors that older adults, as users of eHealth, consider important to embrace the technology [10].

In the process of expanding the use of new technologies, many studies have been conducted to find how users accept specific technologies. The 2 representative models are the Technology Acceptance Model (TAM), which posits that both perceived usefulness and perceived ease of use affect the user's attitude and behavioral intention [11], and the Unified Theory of Acceptance and Use of Technology (UTAUT), which was designed as a synthesis of 8 major technology acceptance models [12]. There are many studies on eHealth acceptance for older adults that extend, transform, or combine these 2 models to identify acceptance factors [13-16]. However, neither is it easy to find evidence for an appropriate acceptance factor model, according to the conditions of the research nor is it easy to construct an optimized acceptance model.

For use in future studies in this field, this study extracts the acceptance factors from studies on eHealth for older people over the past 11 years through a systematic review. After that, we propose an integrated acceptance framework (IAF) that groups the extracted acceptance factors through thematic analysis. Then, the criteria for evaluating the evidence for each factor incorporated into the IAF are provided. Finally, the proposed IAF is analyzed according to the detailed conditions. This study proceeds in 4 steps, from data extraction for the IAF to analysis for the IAF application (Figure 1). One of the main aspects that differentiate IAF from TAM and UTAUT is the presentation of a wider range of factors and dimensions based on acceptance factors that have been covered in research over the past 11 years. Moreover, although TAM and UTAUT are generic acceptance frameworks, IAF is intended to be a framework specifically tailored for a concrete technology (eHealth) and a particular population (older people). In this way, it is expected to be a more useful tool for highlighting the potential barriers and facilitators when planning a new adoption scenario.

Figure 1. The 4-step study model.



Objective and Process

This study can contribute to eHealth research and industry in the following three ways: (1) the development of an IAF, which comprises acceptance factors grouped in dimensions for the analysis of eHealth acceptance in old age that emerge from the analysis of existing evidence; (2) a metric for the assessment of the quality of the evidence found in the systematic review study that allows for the normalization and integration of the evidence on the impact of acceptance factors across different and diverse studies; and (3) stratification analyses of the IAF application according to the participant's health status and verification time and analysis of the evolution of the factors through the years.

Methods

Overview

We conducted a 4-step study (Table 1). First, the primary studies were selected through a systematic review, and the acceptance factors and basic data for analysis were extracted from the selected studies. Second, the extracted factors were grouped through thematic analysis, and a framework for acceptance was proposed. Third, the metrics for quality assurance were defined to evaluate the weights of the factors addressed in the study and apply them to the framework. Finally, we applied the resulting framework for different scenarios—first, according to the health, and second, according to the verification time—resulting in particularized IAFs. Moreover, we analyzed the changes in these factors over the years.

Table 1. Step-by-step study agenda.

Step and agenda	Description
Systematic review	
Which research was selected through a systematic review, which research methods were used for each study, who were the participants, and for which technologies were the acceptance factors studied?	This allows a comprehensive review of research methods and research distribution of selected studies.
Thematic analysis	
What are acceptance factors verified through each study, and can the factors be grouped by thematic analysis to present an IAF[a]?	A quick overview of the selected studies indicates that similar elements are considered in different studies under different terms and with different levels of abstraction. The need to generate an IAF that would emerge from a thematic analysis of the collection of all acceptance elements mentioned in each study, thus grouping similar elements and providing a dimensional classification of acceptance factors, is anticipated.
Evaluation of evidence quality	
Can the importance of each acceptance factor be assessed by combining the evidence provided in different studies?	Given the high variability in the research methods used in the various studies and in the size and characteristics of the participants, the need to establish a metric that assesses the quality of the evidence provided by each study is anticipated. In addition, it is possible to compute a weighted combination of the importance of acceptance factors proposed in the selected studies.
Influential factor analysis	
Stratification analysis for the IAF	
IAF by health status	Through this analysis, it is possible to compare the acceptance factors studied in the group of healthy older adults with the acceptance factors studied in the group of older adults with diseases.
IAF by verification time	The relevant acceptance factors of preadoption (before installation) and those of postadoption (after installation or after use) will be compared and analyzed.
Evolution of factors along the years	The analysis of whether there has been any change in acceptance factors over time, considering the rapidly developing eHealth technology and its growing adoption, is a goal of this research.

[a]IAF: integrated acceptance framework.

Step 1: Systematic Review

This study selected and analyzed the studies according to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. The study scope was defined by the population, intervention, comparison, and outcomes (PICO) model, and accordingly, the search scope and research questions were defined. In the screening stage, 3 reviewers collaborated using the Covidence tool (Veritas Health Innovation Ltd). During the process of extracting and organizing data, a Microsoft Excel spreadsheet was used and shared for collaboration through Microsoft Teams.

The PICO model was applied to define the research scope and search strings as follows:

- Population: older adults who have the possibility of using eHealth
- Intervention: eHealth technology (including mHealth, telemedicine, and telehealth) that older users may experience or think about
- Comparison: comparison among the participants' conditions, comparison based on verification time, and comparison of changes in acceptance factors by year
- Outcome: extraction and consolidation of acceptance factors and their impact on the adoption of eHealth services for older adults

Search strings were defined as combinations that can retrieve as many related studies as possible, with consideration given to PICO. Our final search string was *(ehealth OR telehealth OR mhealth OR uhealth OR health technology OR telemedicine) AND (older OR elderly OR senior) AND (adoption OR acceptance) AND (factors OR barriers OR determinants OR facilitators)*. The search scope was established as article title, abstract, and keywords. The databases used were Web of Science, Scopus, PubMed, IEEE, and MEDLINE. The review was conducted on conference papers or journal articles published during the 11 years from 2010 to 2020.

The set of studies collected through the search was finally selected using the following criteria and quality evaluation questions.

The exclusion criteria were as follows:

- Articles not written in English
- Articles that did not directly use the terms *acceptance* and *health technology* or related terms in the title, abstract, or entire text

- Studies that discuss eHealth adoption factors but not for older users
- Meta-analysis reviews the same subject

The quality evaluation questions were as follows:

- Are the influential factors clearly defined?
- Is the empirical evidence presented?
- Are the ages of the participants clearly stated (mean age of ≥60 years)?
- In the case of quantitative research, is the number of participant responses sufficient?
- In the case of qualitative research, has there been sufficient discussion of acceptance factors?

The data to be extracted from each study were defined according to the agendas in each step, as shown in Textbox 1. The data were extracted and organized in step 1, and the extracted data from each agenda were used in each step.

Textbox 1. Data definitions for extraction.

Year of publication

- Year the study was published

Country

- Countries subjected to study

Participants' mean age

- Average age of participants

Verification time

- When the acceptance factors are verified

Study method

- The methods used to study the acceptance factors

Technology

- Health-related technologies in studies

Theory

- Theories on which the study is based

Participants' condition

- Participants' health status or recruitment conditions

Factor, barrier, or facilitator

- Factors, barriers, and facilitators tested

Result

- Research results and insights

Step 2: Thematic Analysis

The high number and diversity of factors, together with the variety of research methods, make it difficult to collect existing evidence and reach meaningful conclusions about the factors that really have an impact on the acceptance and adoption of these technologies. To overcome this, we formed an integrated framework for eHealth acceptance factors in older adults. All factors extracted from the selected studies by the systematic review process were defined as either positive or negative. Then, a thematic analysis process was conducted with the goal of identifying, analyzing, and interpreting patterns of meaning (or *themes*) within the set of original acceptance factors. We grouped them according to commonality in the meaning of the original acceptance factors in a bottom-up fashion. The 3 authors jointly analyzed and classified the acceptance factors through Microsoft Teams, and they reviewed and discussed each article to understand the meaning of the factors used in that article. Concretely, the first author conducted an initial thematic analysis after data extraction. Later, the other 2 authors participated in several consensus meetings. The other authors are senior researchers in two complementary disciplines: the first is a professor of computer science with a degree in psychology and extensive experience in acceptance models, and the second is a professor of biomedical engineering with wide experience in eHealth and older adults using technologies. After extracting and defining factors for 3 weeks, grouping was conducted according to themes for an additional 2 weeks.

Step 3: Evaluation of Quality of Evidence

To assess the impact of the factors, we evaluated and reflected the quality of evidence in each study beyond the frequency of the factors used in the studies. For a systematic review, there is a grading of recommendations, assessment, development, and evaluation (GRADE) method that evaluates the quality of the evidence for each outcome by applying a set of evaluation criteria [17]. With GRADE, the quality of the evidence is evaluated according to the research method, as well as the risk of bias, inconsistency, indirectness, imprecision, and large magnitude of effect. However, GRADE focuses on the results of the study rather than evaluating the overall quality of the research and is mainly targeted for experiments related to health care. Consequently, we found it difficult to apply GRADE in our review. Thus, we only took the research methods criteria considered in GRADE and other studies [17,18], and we felt it was necessary to add some new criteria that could assess the quality of evidence of the selected studies. One of the outcomes of this research is a metric that defines a set of relevant evaluation criteria.

The impact of the acceptance factors was analyzed by calculating the evidence quality score of each study according to this metric and deriving from it the weight of the acceptance factors studied.

In this way, the proposed IAF was enhanced by reflecting the impact of acceptance factors.

Step 4: Influential Factor Analysis of the IAF

We analyzed the IAF that resulted from the research to better understand the relative importance of the factors according to different conditions. To do this, a stratification analysis [19] was applied that allowed the classification and analysis of the factors according to the conditions. First, the factors that depend on the health condition of the participant were analyzed; then, the analysis was repeated according to the verification time (preadoption and postadoption). In addition, the IAF was analyzed by year to examine the evolution of factors over the years.

Results

IAF Steps

Step 1: Systematic Review

According to the PRISMA guidelines, of the 731 studies retrieved from the databases, after excluding duplicates, 168 (23%) studies were screened. Of the 168 studies, after excluding 94 (55.9%) studies that were considered irrelevant, a total of 74 (44%) studies were reviewed for full text during the eligibility phase. Figure 2 shows the PRISMA flowchart of the study selection process, where 58% (43/74) of the articles were finally selected. The selected studies clearly identified acceptance factors for eHealth or health technology for the older population.

Multimedia Appendix 1 [13-16,20-58] lists the 43 studies and their corresponding basic data (see Textbox 1 for definitions). Of the 43 selected publications, 27 (63%) reported quantitative studies, 10 (23%) reported qualitative studies, and 6 (14%) reported mixed methods studies. Only 9% (4/43) of them were longitudinal studies, which observed and analyzed the same group for a long period, and their continuous observation period ranged from 3 months to 2 years. Quantitative research was mainly conducted as a survey in the form of mail or web-based questionnaires, and because of the characteristics of old age, there were also studies conducted through a face-to-face survey with explanations about the research. For qualitative research, in-depth interviews and focus group interviews were conducted at a similar rate.

Participants' health conditions for each investigation were also identified. Of the 43 studies included in this research, 26 (60%) studies only considered participants without pre-existing disease conditions, followed by 9 (21%) studies that included participants with chronic diseases. Of the 43 studies, there was 1 (2%) study that compared healthy participants to participants with chronic diseases, and 2 (5%) studies compared healthy participants to participants with heart disease.

Figure 3 displays the distribution of health technologies that need to be studied. mHealth was the most common classification, with 28% (12/43) of studies, followed by eHealth with 23% (10/43) of studies, which dealt with general and integrative health technology.

Most studies were based on existing technology acceptance theories, 40% (17/43) of studies were based on TAM, and 21% (9/43) of studies were based on UTAUT.

The distributions of the selected 43 studies, according to country and year, are shown in Multimedia Appendix 2. Excluding an anomalous decrease in 2018, the trend is a growing number of papers, with recent studies in 2019 and 2020 accounting for a large proportion. The largest number of studies was conducted in the United States, followed by studies conducted in China.

Figure 2. PRISMA (Preferred Reporting Item for Systematic Reviews and Meta-Analyses) flowchart for study selection.

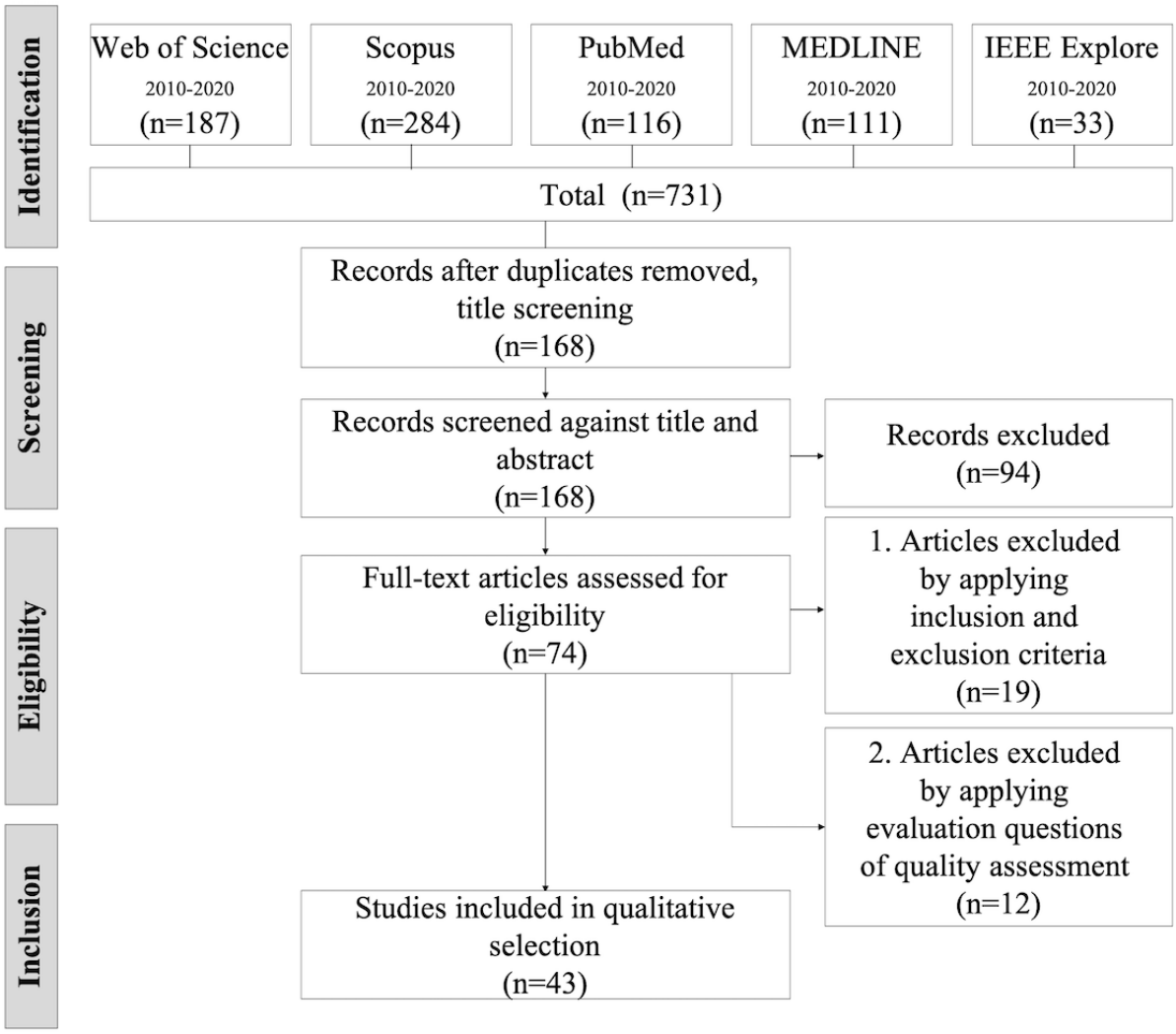
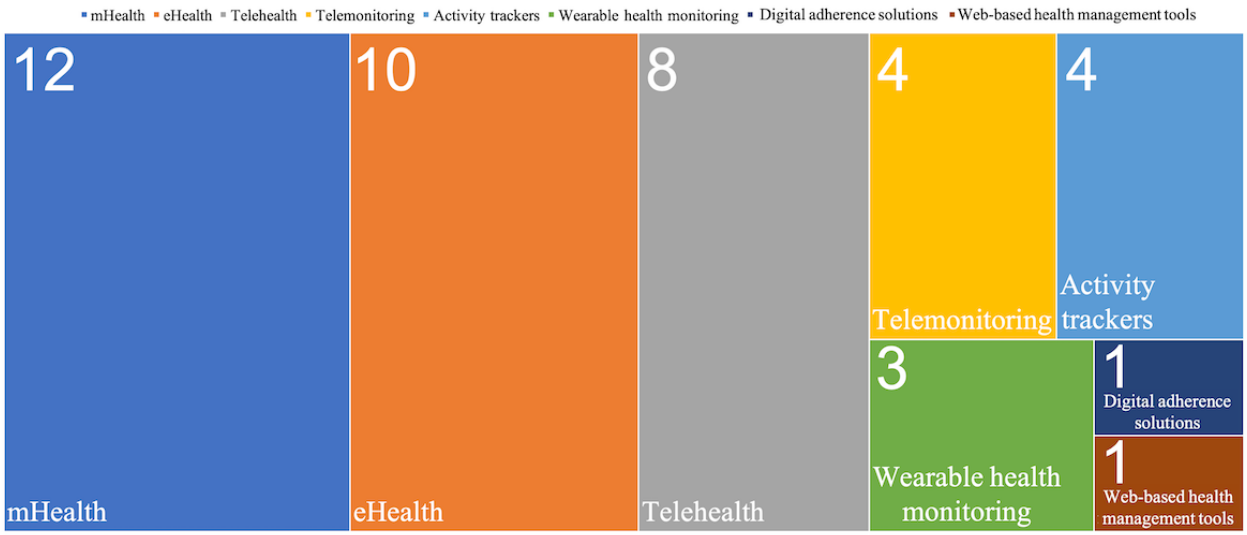


Figure 3. Target technology. mHealth: mobile health.



Step 2: Thematic Analysis

A total of 105 acceptance factors were extracted from the studies. The 105×43 matrix relating acceptance factors to the studies was too sparse, and the absolute frequency of acceptance factors across the studies was too low. It was obvious that this could not be taken as the basis for the combination of evidence. Therefore, a thematic analysis process was conducted with the

goal of identifying and analyzing the patterns of themes within the set of original acceptance factors. By grouping according to commonality in the meaning of original acceptance factors in a bottom-up fashion, the resulting IAF comprised 23 representative acceptance factors or themes, which were categorized into five dimensions: (1) personal, (2) user–technology relational, (3) technological, (4) service relational, and (5) environmental (Figure 4). The details of the 23 final acceptance factors, their corresponding elements (rephrasing the original acceptance factors), the frequency with which they are analyzed across the 43 studies, and the type of influence they have been found to exert on the decision to use eHealth (positive or negative), are included in Table 2.

The *personal* dimension comprises a total of five factors related to the user: (1) *personal characteristics*, which comprise an individual's basic profile; (2) *personal condition* to reflect an individual's health status or activity level; (3) *personal capabilities* to know eHealth acceptance capacities; (4) *personality and attitude*, which considers all personal traits, beliefs, and attitudes that can have an impact on the adoption of eHealth technology; and (5) *preferences*, which reflect personal inclinations for health care.

The *user–technology relational* dimension comprises five factors that lie in the intersection between the user and technology: (1) *how technology addresses user needs/characteristics* to consider the degree of matching between the technology and the real needs of the user; (2) *experience with technology* to take into account previous experience with other technologies; (3) *perceived usefulness of technology*; (4) *perceived ease of technology*; and (5) *attitude toward technology*, which groups concerns and feelings that the user has toward the technology.

The *technological* dimension comprises six factors related to the technology: (1) *features/functions* of technology, (2) *quality of technology and device*, (3) *usability* of technology, (4) *hedonic motivation* of technology, (5) *automaticity*, and (6) *benefits for users* of technology use.

The *service-related* dimension comprises a total of five factors that consider the service aspects in the adoption of eHealth technology: (1) *support for use*, (2) *cost for eHealth service*, (3) *service quality*, (4) *organizational factors* related to service operation, and (5) *alignment with government policies*.

Finally, the *environmental* dimension comprises two elements that address the context in which the user will adopt the technology: (1) *physical environments*, such as distance from hospitals and internet accessibility, and (2) *social influence*, which is influenced by people around older users.

The factors of *personality and attitude* in the *personal* dimension included most of the elements. These elements correspond to personal dispositions, personality traits, or attitudes that can influence the acceptance of new things or changes in the way they deal with health. The factor of *attitude toward technology* in the *user–technology relational* dimension comprises elements that mostly affect the acceptance of eHealth technology in a negative way, such as technology anxiety, privacy concerns, and security concerns.

The *personal characteristics* factor, the element of chronic health condition or health status in the *personal condition* factor, and the element of the degree of satisfaction with existing medical services in the *how technology addresses user needs/characteristics* factor have shown both positive and negative effects on eHealth acceptance, depending on the study.

Thus, we propose an IAF that comprises these 23 grouped factors and 5 dimensions to be used as a reference in future research on eHealth acceptance.

Figure 4. The 5 dimensions for acceptance factors.

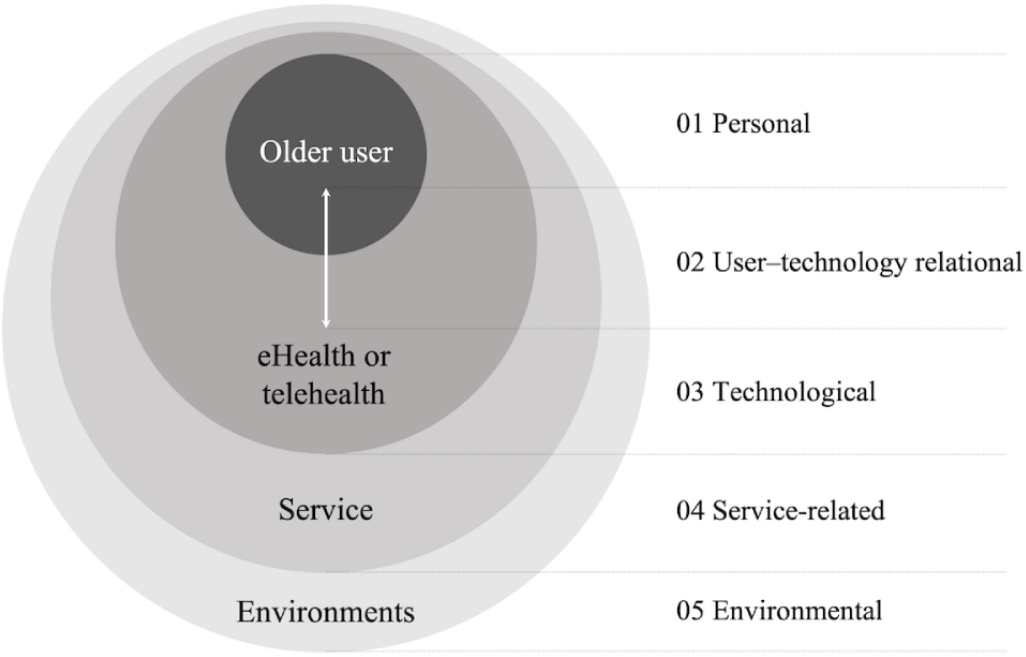


Table 2. Proposed integrated acceptance framework with 23 acceptance factors and their elements (N=43).

Factor and element	Frequency, n (%)	Influence	
		Positive	Negative
Personal			
Personal characteristics			
Age	6 (14)	✓	✓
Gender	5 (12)	✓	✓
Educational background	4 (9)	✓	✓
Lifestyle and residence type	4 (9)	✓	✓
Income	2 (5)	✓	✓
Work status	2 (5)	✓	✓
Adequate financial status	1 (2)	✓	
Geographical location	1 (2)	✓	✓
Health knowledge	1 (2)	✓	
Personal condition			
Chronic health condition or health status	8 (19)	✓	✓
High activity level	1 (2)	✓	
Independence	1 (2)	✓	
Personal capabilities			
Self-efficacy or competence	9 (21)	✓	
Decreased physiological or cognitive capability	6 (14)		✓
Participation	1 (2)	✓	
Personality and attitude			
Concerns about risk	5 (12)		✓
Conversion readiness or personal innovativeness	4 (9)	✓	
Resistance to change	2 (5)		✓
Personal proactivity	2 (5)	✓	
Sense of control	2 (5)	✓	
Confidence in control of health	1 (2)	✓	
Overanxiety about health	1 (2)		✓
Perceived social risk	1 (2)		✓
Need for cognitive closure	1 (2)		✓
Willingness to take a chance	1 (2)	✓	
Ability to take advantage of opportunities	1 (2)	✓	
Self-esteem	1 (2)	✓	
Self-confidence	1 (2)	✓	
Reluctance to rely on a machine	1 (2)		✓
Preferences			
Preference for face-to-face contact	3 (7)		✓
User–technology relational			
How technology addresses user needs or characteristics			
Lack of needs	4 (9)		✓
Degree of satisfaction with existing medical service	2 (5)	✓	✓
Insufficient contents or functions	1 (2)		✓

Factor and element	Frequency, n (%)	Influence	
		Positive	Negative
Needs are already addressed by caregiver	1 (2)		✓
Desire for ownership of and access to medical information	1 (2)	✓	
Information or system feature overload	1 (2)		✓
Health care needs	1 (2)	✓	
Experience with technology (literacy)			
Lack of information and awareness	7 (16)		✓
Prior experience with technology	6 (14)	✓	
eHealth literacy	4 (9)	✓	
Poor eHealth experience	3 (7)		✓
Frequency of internet use	1 (2)	✓	
Perceived usefulness of technology			
Perceived usefulness	20 (47)	✓	
Performance expectation	8 (19)	✓	
Perceived security	7 (16)	✓	
Perceived compatibility	4 (9)	✓	
Perceived ubiquity	1 (2)	✓	
Perceived relative advantage	1 (2)	✓	
Perceived ease of technology			
Perceived ease of use	18 (42)	✓	
Difficulty with new technology	8 (19)		✓
Effort expectation	8 (19)	✓	
Perceived complexity of technology	2 (5)		✓
Amount of perceived effort	1 (2)		✓
Attitude toward technology			
Technology anxiety	13 (30)		✓
Privacy concerns	8 (19)		✓
Lack of interest	4 (9)		✓
Security concerns	4 (9)		✓
Lack of trust in service	4 (9)		✓
Trust in service	3 (7)	✓	
Negative feeling about constant monitoring	1 (2)		✓
Technological			
Features or functions			
Track vital signs or monitor my information	3 (7)	✓	
Functions to help existing health care services	2 (5)	✓	
Monitor health trends	1 (2)	✓	
Quality of technology and device			
Technology instability	4 (9)		✓
Convenience	2 (5)	✓	
Physical comfort (wearable)	1 (2)	✓	
Usability			
Insufficient user-friendliness	6 (14)		✓

Factor and element	Frequency, n (%)	Influence	
		Positive	Negative
Learning difficulty of new technology	2 (5)		✓
Lack of instructions	2 (5)		✓
Esthetics	1 (2)	✓	
Helpful instructions	1 (2)	✓	
Hedonic motivation			
Hedonistic motivation	1 (2)	✓	
Automaticity			
Using it everyday	1 (2)	✓	
Using a variety of functions	1 (2)	✓	
Habit	1 (2)	✓	
Benefits for user			
Share data with someone	3 (7)	✓	
Digital solutions that remove personal barriers	2 (5)	✓	
Medical records in one place	2 (5)	✓	
Observation of changes after use	2 (5)	✓	
Portable personal records	1 (2)	✓	
Prevention of unnecessary tests or medical accidents	1 (2)	✓	
Service related			
Support for use			
Technical support	6 (14)	✓	
Support from people around me	6 (14)	✓	
Peer support	6 (14)	✓	
Adequate training	4 (9)	✓	
Intergenerational support	4 (9)	✓	
Support from service	4 (9)	✓	
Hospital support	2 (5)	✓	
Not enough support for technology use	1 (2)		✓
Cost for eHealth service			
Cost burden	9 (21)		✓
Service affordability	4 (9)	✓	
Service availability	3 (7)	✓	
Price value	2 (5)	✓	
Service quality			
Information quality or service quality	2 (5)	✓	
Organizational factors			
Care assistance center linked to service	1 (2)	✓	
Improvement of health care interactions	1 (2)	✓	
Provided in parallel with existing direct visits	1 (2)	✓	
Alignment with government policies			
Government policy	1 (2)	✓	
Environmental			
Physical environments			

Factor and element	Frequency, n (%)	Influence	
		Positive	Negative
Internet connection instability	3 (7)		✓
Distance to hospital	1 (2)	✓	
Social influence			
Social norms or subjective norm	12 (28)	✓	
Physician's recommendation	6 (14)	✓	
Recommendation from people around me	5 (12)	✓	
Family recommendation	1 (2)	✓	
Close people's eHealth readiness	1 (2)	✓	

Step 3: Evaluation of Quality of Evidence

Overview

The assessment results should quantify the reliability of the findings on acceptance factors discussed in each study and whether their conclusions can be confidently applied to future related studies. The proposed metric is based on three criteria:

1. Reliable methodology
2. Participant's experience with the specific target technology
3. Research and publication year

The score for each of the 3 criteria ranged from 1 to 4: very low=1, low=2, moderate=3, and high=4. As studies with *very low (1 point)* quality by these criteria have already been excluded through the quality assessment of the full text, the selected studies received scores ranging from 2 to 4 points.

Reliable Methodology

The quality of evidence increases when the applied research methods can provide high internal and external validity. It is considered that the validity of quantitative studies strongly depends on the number of participants. We also considered that mixed approaches, in which quantitative and qualitative methods are combined, tend to have higher validity than single methods. Finally, the clarity and reliability of the analysis method also have an influence on validity. The rules applied for the assessment were as follows:

- High (4 points): studies with multidimensional approaches that applied a mixed or longitudinal study
- Moderate (3 points): quantitative studies with sufficient participants and clear analysis methods; qualitative studies following a reliable analysis method
- Low (2 points): quantitative research with <100 participants; qualitative research that did not mention a clear analysis method

Participant's Experience of the Specific Target Technology

This criterion evaluates whether the target technology was clearly explained to or experienced by the participants before discussion and investigation. With a high degree of understanding of the technology being studied, participants could express their intentions more accurately. Otherwise, their answers could be biased by misunderstandings or prejudices. The rules applied for the assessment were as follows:

- High (4 points): The subject of investigation clearly recognized the target technology through a prototype or demonstration video, or the target technology was used for a certain period.
- Moderate (3 points): The technology was presented through text, images, or explanations from the investigator. Alternately, participants had an indirect understanding of the target technology based on their previous experiences with other technologies.
- Low (2 points): The method for the presentation of the target technology to the participants was not mentioned in the article.

Publication Year

Recently, eHealth has been developing at a faster rate. In this context, additional points were applied to recent research in consideration of the fact that acceptance factors can be affected by recent advancements in technology and infrastructure. Moreover, the adoption of eHealth technologies has been increasing over time, possibly leading to a change in the influence exerted by some acceptance factors.

- High (4 points): studies from 2017 to 2020
- Moderate (3 points): studies from 2014 to 2016
- Low (2 points): studies from 2010 to 2013

Quality of Evidence

We followed four steps to calculate the influence of the acceptance factors:

1. Each study obtained a quality score according to the defined metric, with the results ranging from a maximum of 12 points to a minimum of 8 points.
2. This score was normalized by transforming a perfect score of 12 into 1 and a score of 8 into 0.67.
3. The occurrence of each factor in a study was represented by the normalized score for the corresponding study.
4. The values for all occurrences of a factor were totaled.

Through this process, the real influence of the acceptance factors was measured in a more reliable way than just by considering the absolute frequency. Figure 5 shows the impact order of the factors, reflecting the quality of the supporting evidence. The Pareto chart allows for the selection of important elements in the 80% criterion by cumulative impact [59]. Applying the Pareto chart, 10 factors can be distinguished among the 23

factors in the IAF, as shown in the chart (Figure 5): *perceived usefulness of technology, attitude toward technology, perceived ease of technology, support for use, personal characteristics, social influence, personality and attitude, experience with technology, cost of eHealth service, and personal capabilities*. Together, these account for 80% of the found evidence of impact.

Figure 6 depicts the impact of the factors classified according to the 5 dimensions of the IAF.

The weights of the 5 dimensions were calculated as the sum of the weights of the factors corresponding to each dimension. As shown in Figure 6, as the weights of the factors constituting each dimension are different, the weights of each dimension are also different.

The acceptance factors in the *user–technology relational* dimension were considered the most important, and *support for use* in the *service-related* dimension, *personal characteristics* in the *personal* dimension, and *social influence* in the *environmental* dimension were also determined to be important. The *technological* dimension was evaluated to be less important than the other dimensions.

These impact scores for the factors in the IAF come to complete the framework and define the relative importance of each factor.

Figure 5. Factor impact based on quality assessment of evidence on integrated acceptance framework.

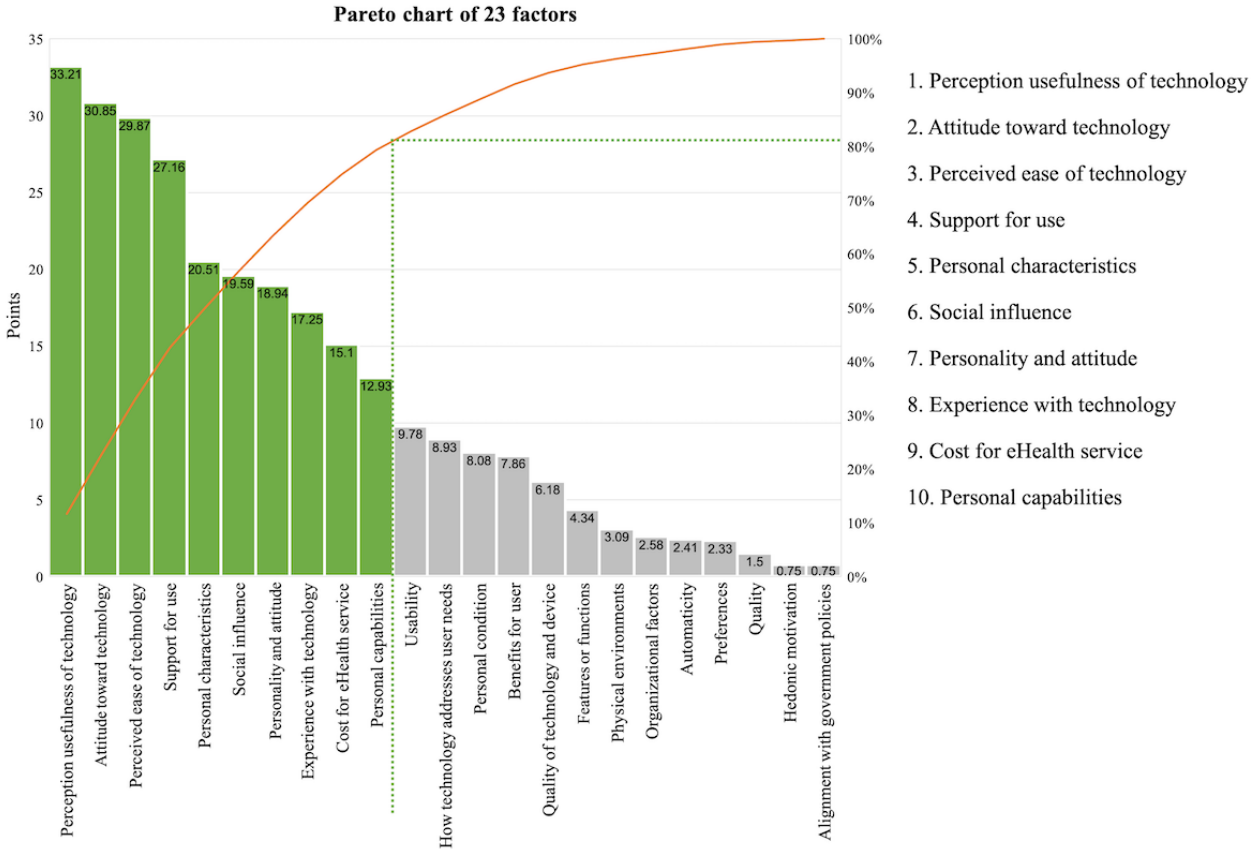
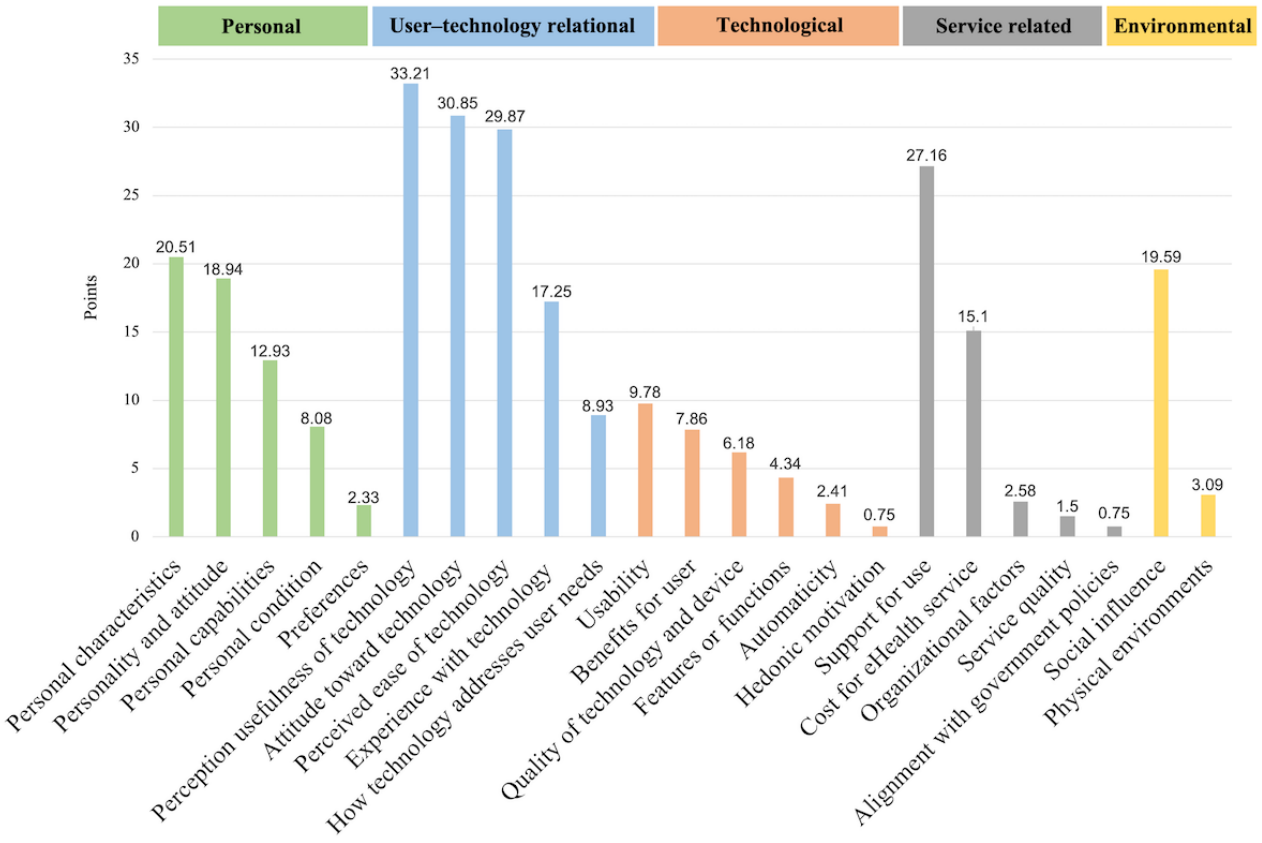


Figure 6. Factor impact on each dimension of integrated acceptance framework.



Influential Factor Analysis

Overview

This section presents the results of the fourth step. The IAF proposed through the previous 3 steps was analyzed under 3 conditions. The following subsections describe the optimized IAF for each specific situation.

Stratification Analysis of IAF by Health Status

As previously stated in step 1, of the 43 studies, 14 (33%) involved older adults with specific diseases (eg, chronic diseases and heart disease), and 26 (60%) included older adults without disease conditions. In addition, 7% (3/43) of studies [46,47,55] compared the healthy group and the group with diseases. It may be a risk to conclude that the participants without a disease condition are healthy; however, it is possible to determine that they are healthy compared with the group with a specific disease. Therefore, for the comparison based on participants' conditions, the studies with older people without disease conditions were classified as the healthy group, whereas studies considering older people with diseases were classified as the unhealthy group, and the differences in the acceptance factors of these 2 groups were analyzed (Figure 7).

After applying the quality score of each study's evidence, the impact of factors was calculated as a percentage, as the number of studies in each group was different. As a result, the most important factor in studies in the healthy group is the *perceived usefulness of technology*. In addition, the following factors were considered more important for the healthy group than for the unhealthy group: *support for use*, *social influence*, *perception of technology*, *personality and attitude*, and *personal characteristics*. However, the *preferences* factor in the IAF was not considered at all in the studies for this group. On the other hand, in the case of the unhealthy group, the most important acceptance factor was the *perceived ease of technology*. The factors *physical environment*, *cost of eHealth service*, *usability*, *quality of technology and device*, *perceived ease of technology*, and *personal capabilities* were considered more important for this group than for the healthy group. In this group, *personal condition*, *hedonic motivation*, *automaticity*, and *alignment with government policies* were not considered as acceptance factors. It was confirmed that the 2 groups showed different eHealth acceptance factor patterns. These results can be interpreted as evidence of variance in acceptance factors according to participants' health conditions.

Through the Pareto chart (Figure 8), nine acceptance factors for the healthy group IAF were identified: *perceived usefulness of technology*, *perceived ease of technology*, *support for use*, *attitude toward technology*, *social influence*, *personality and attitude*, *personal capabilities*, *experience with technology*, and *personal characteristics*. In addition, the following factors for studies on participants with diseases were found: *perceived ease of technology*, *perceived usefulness of technology*, *attitude toward technology*, *experience with technology*, *usability*, *cost of eHealth service*, *personal capabilities*, *how technology addresses user needs*, *quality of technology and device*, *personality and attitude*, *and support for use*.

Figure 7. Comparison of the acceptance factor impact by participant health status.

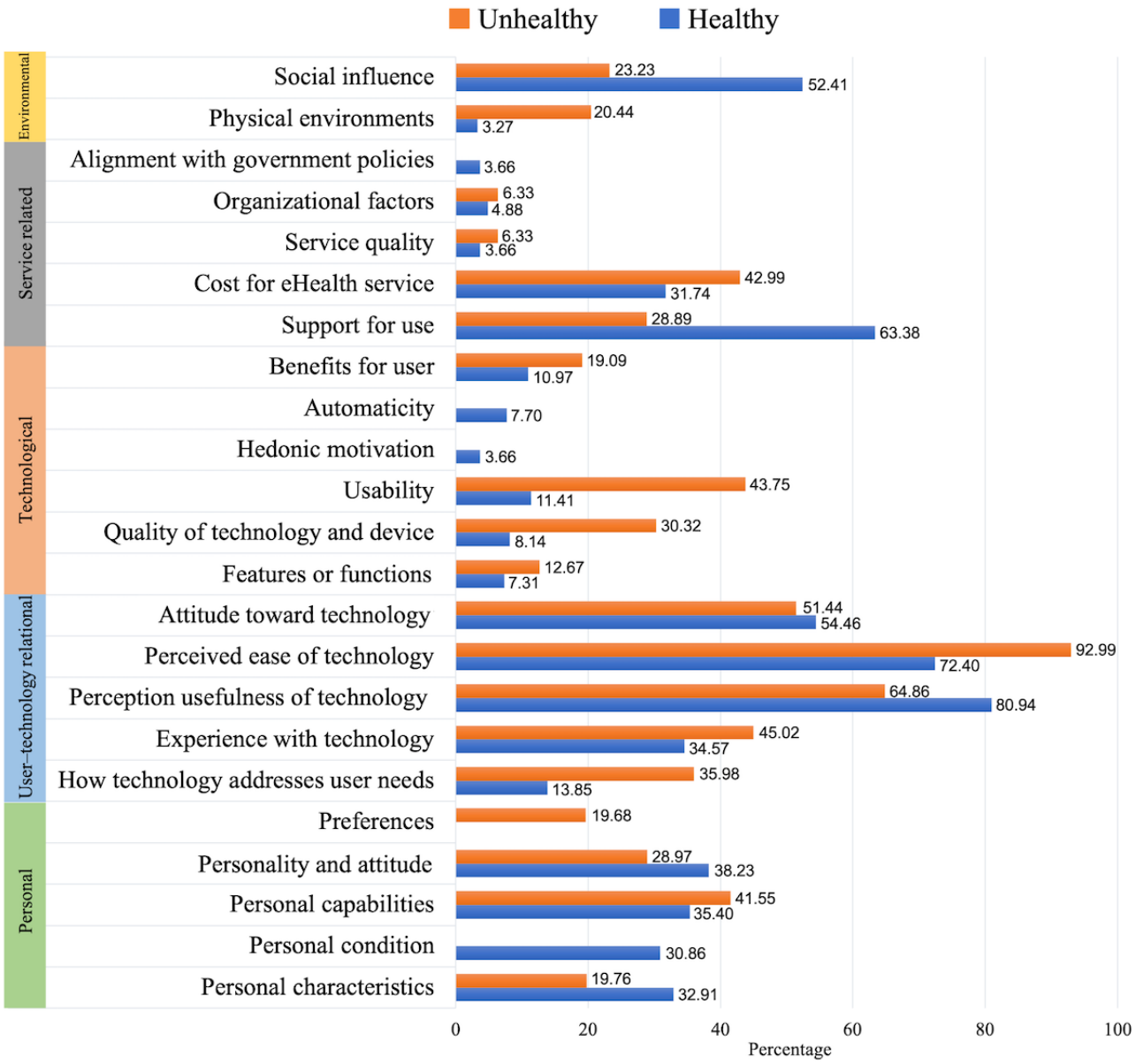
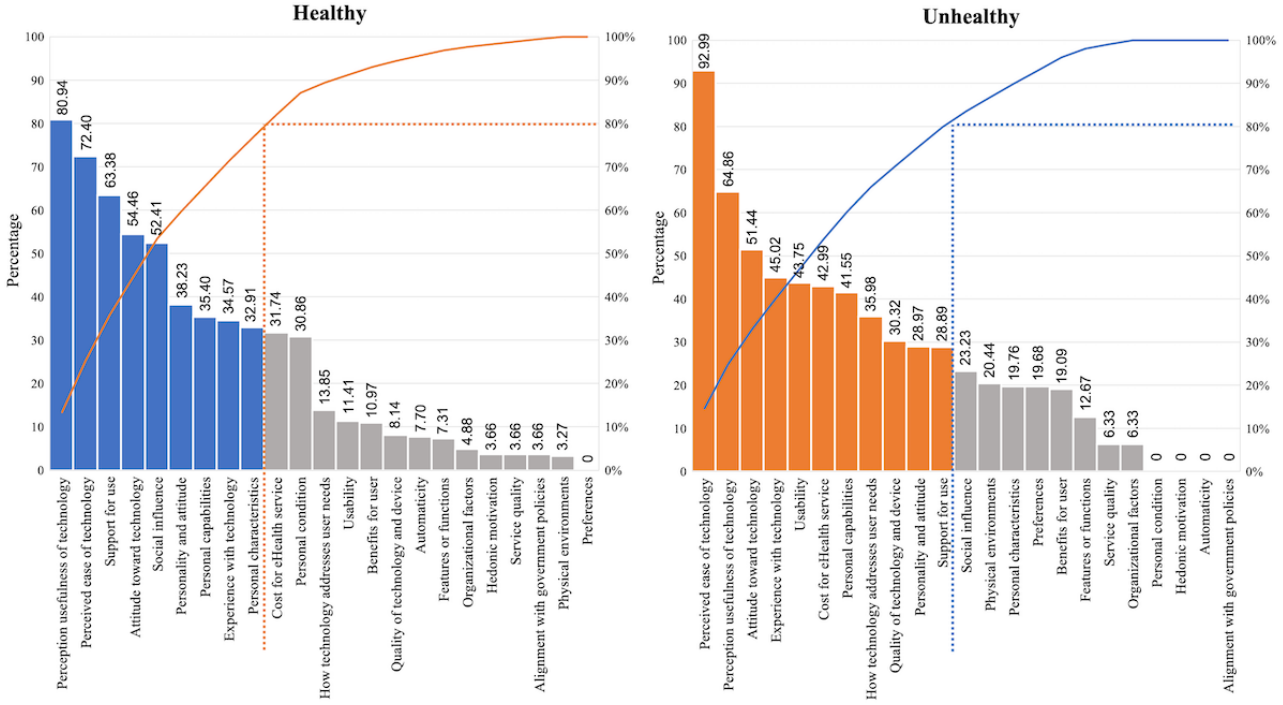


Figure 8. Pareto chart for integrated acceptance framework by participant health status.



Stratification Analysis of IAF by Verification Time

We also analyzed the acceptance factors according to the time at which they were verified. In the results for step 1 (Multimedia Appendix 1), the verification time of the studies was divided into three stages: (1) before installation, (2) after installation or at the beginning of use, and (3) after enough use. In this analysis, we decided to compare preadoption and postadoption so that *before installation* was classified as preadoption, and the rest (*after installation/at the beginning of use* and *after enough use*) were classified as postadoption. The 43 studies were divided into 24 (60%) studies related to preadoption, 13 (30%) studies related to postadoption, and 6 (14%) studies that dealt with both situations. The acceptance factors according to the verification time were compared and analyzed, as shown in Figure 9.

The most important factor in the preadoption stage is the *perceived ease of technology*. The factors of *social influence, cost of eHealth service, perception of technology usefulness, personality and attitude*, and *personal characteristics* are more important at this verification time in studies on preadoption than in studies on postadoption. Moreover, the *organizational factors* from the IAF were not considered at all in the preadoption studies. The key factor for postadoption is the *perceived usefulness of technology*. *Organizational factors, support for use, quality of technology and device*, and *personal conditions* are more important acceptance factors in studies on postadoption than in studies on preadoption. At this verification time in the selected studies, the factors of *service quality*, *alignment with government policies*, and *physical environments* were not addressed as acceptance factors. As such, it is confirmed that the degree of influence of the acceptance factors from the IAF differed depending on the verification time.

Applying Pareto (Figure 10), nine acceptance factors for the preadoption phase were identified: *perceived ease of technology, perceived usefulness of technology, attitude toward technology, support for use, personality and attitude, social influence, cost of eHealth service, experience with technology*, and *personal capabilities*. The top 11 factors for the postadoption phase were also identified: *perceived usefulness of technology, perceived ease of technology, attitude toward technology, support for use, quality of technology and device, social influence, cost of eHealth service, personality and attitude, usability, personal conditions, and personal characteristics.*

Figure 9. Comparison of the impact of acceptance factors by verification time.

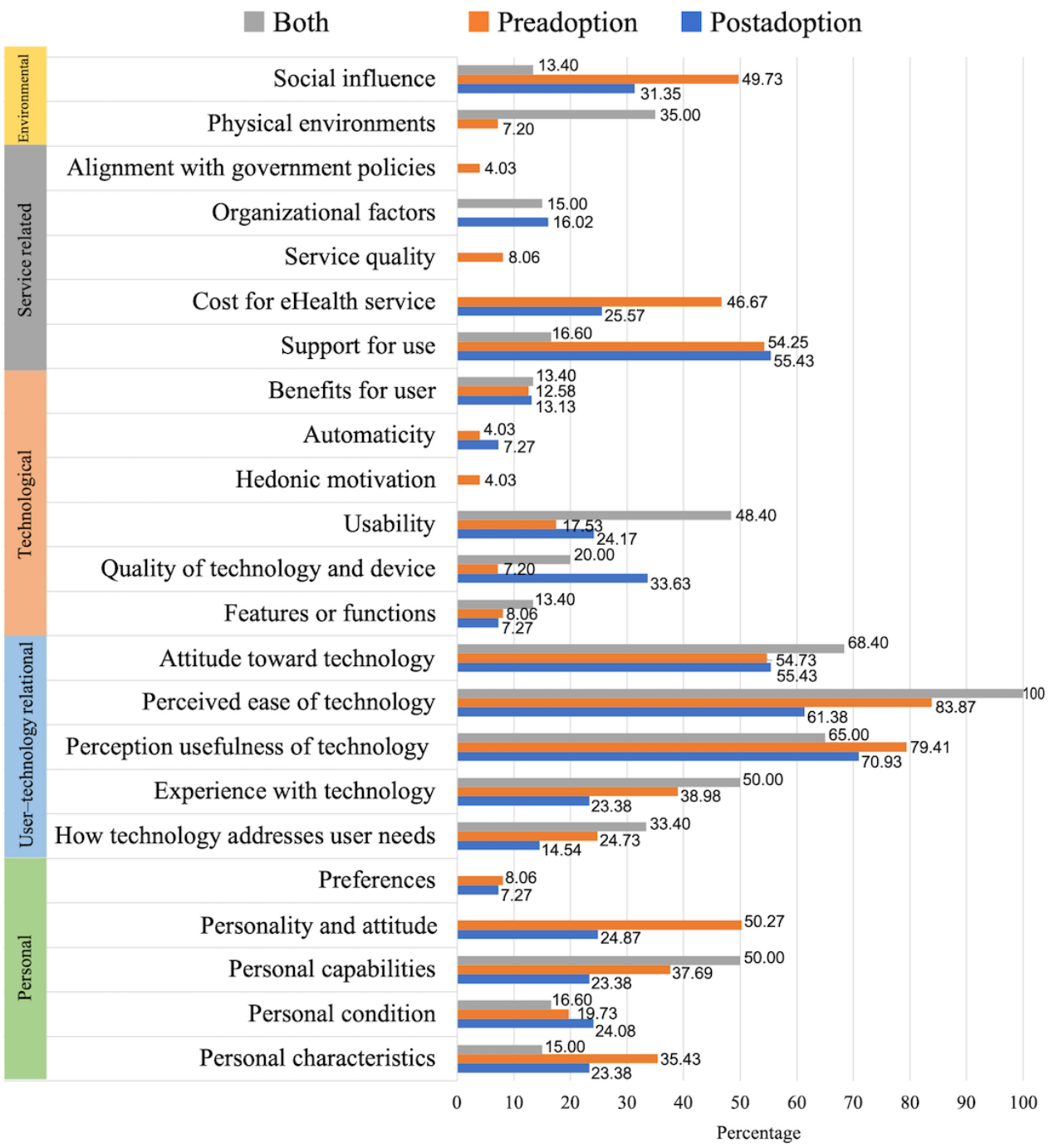
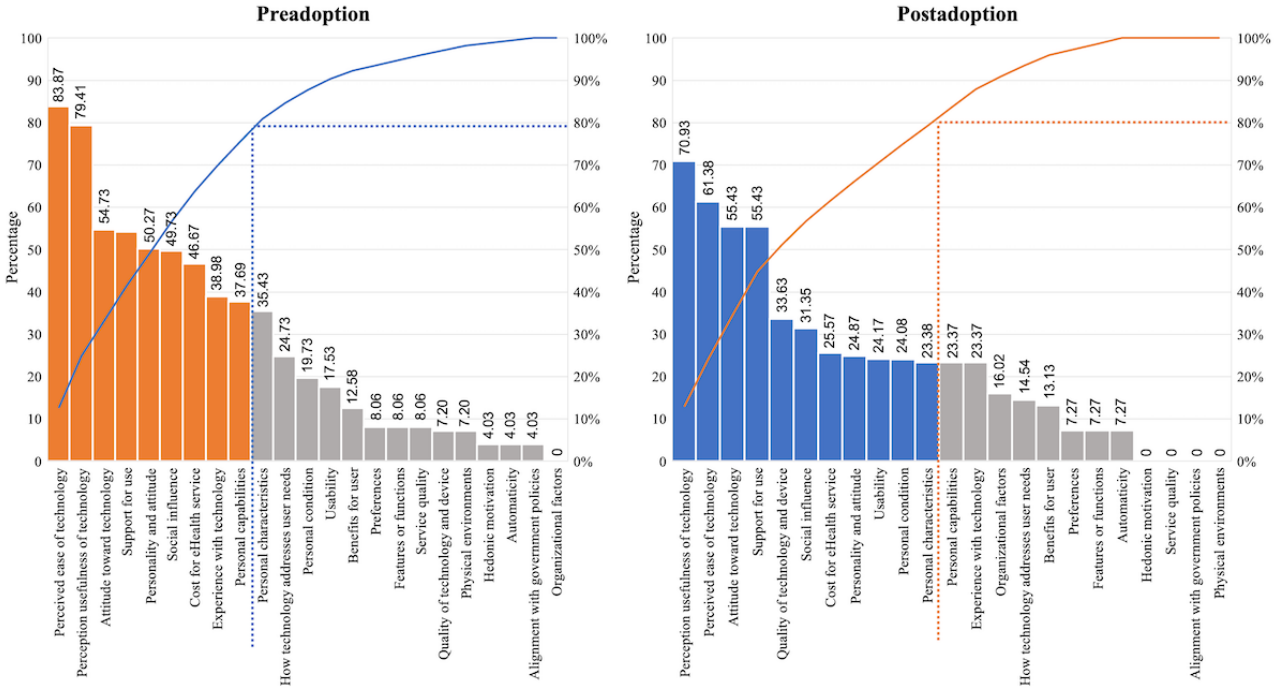


Figure 10. Pareto chart for integrated acceptance framework by verification time.



Analysis on Evolution of Factors Along the Years

As shown in Figure 11, the acceptance factors that become more important with the passage of time are identified by analyzing the changes in the influence of the IAF factors per year. In the *personal* dimension, the pattern change was not very remarkable; however, it shows that the consideration for the *personal capability* factor has recently increased.

The *user–technology relational* dimension is more consistently considered than the other dimensions. In particular, the *perceived ease of technology* and the *perceived usefulness of technology*, which are addressed by the TAM and UTAUT models, are factors that have been steadily considered since the beginning of 2010. In addition, in the *user–technology relational* dimension, the factors of *experience with technology* and *attitude toward technology* have been mentioned since late 2015, and these factors have been considered more recently as well. Although it was difficult to find clear change patterns in the *technological* and *environmental* dimensions, the *support for use* factor in the *service-related* dimension has been identified as a key factor in recent years.

Figure 12 illustrates the changes in the factors within each dimension through a linear graph, along with the accumulated changes in the 5 dimensions per year. The *user–technology relational* dimension has gained importance recently, and the importance of the *service-related* dimension has been steadily rising.

Figure 11. Impact change on integrated acceptance framework by year.

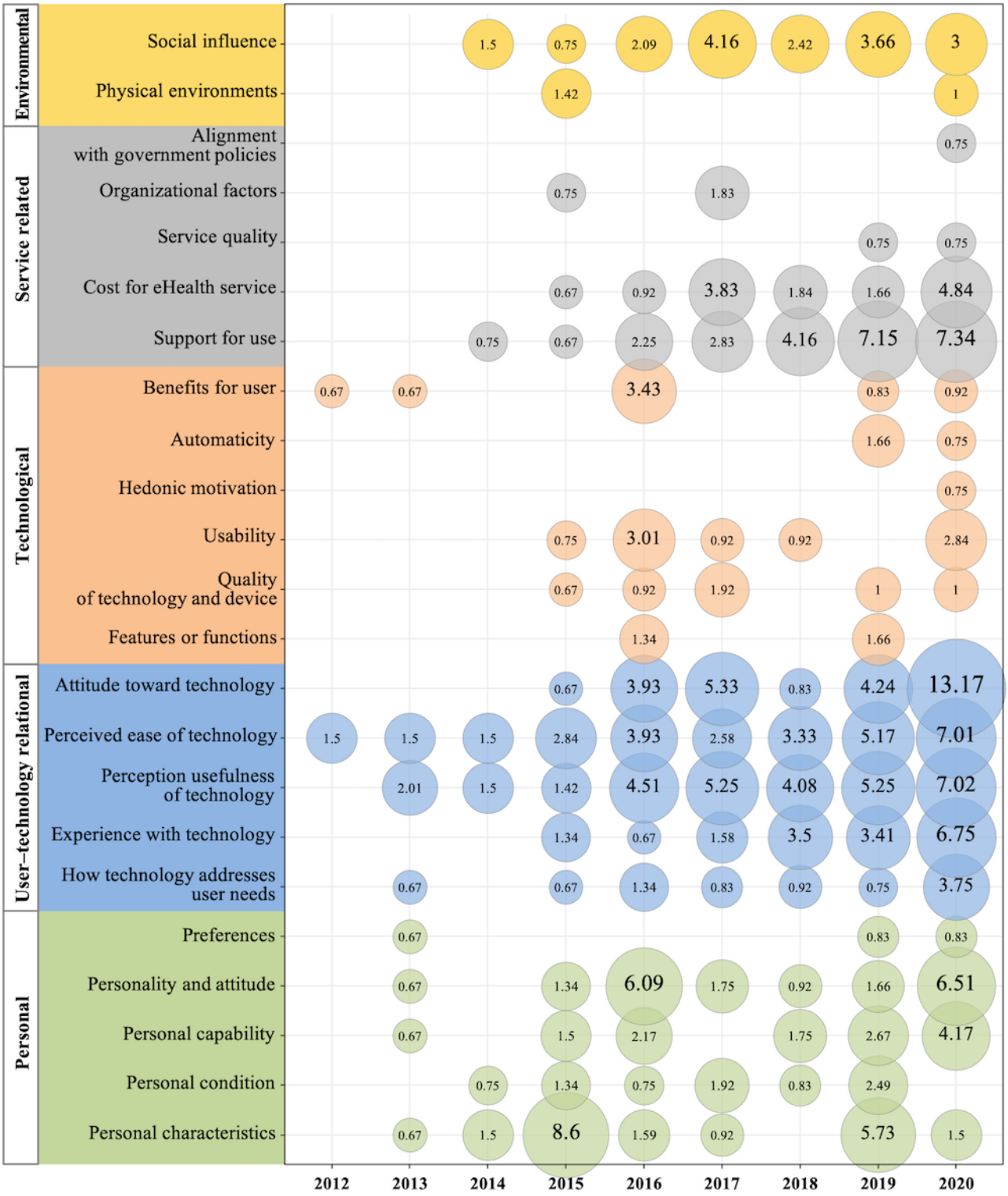
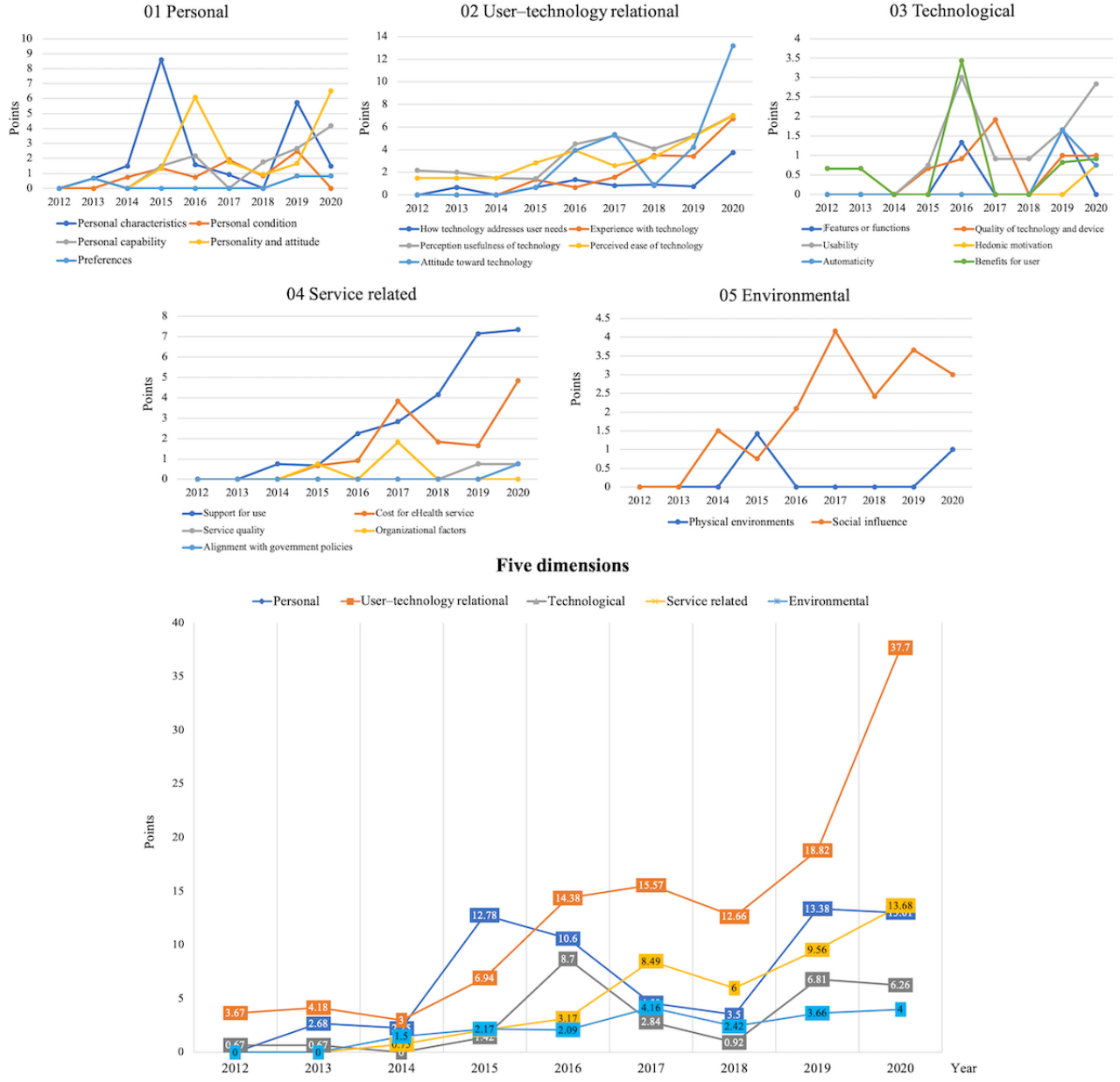


Figure 12. Linear change in the influence of acceptance factors by year.



Discussion

Definition of an IAF

In step 1, we conducted a systematic review to select qualified studies and extract the data as the basis for the IAF. The basic data from the research were extricated, and those became the base material for the formation of the various comparative groups and the assessment of the evidence quality for the selected studies. A previous integrative review on the adoption of telehealth in older age [60] extracted facilitators and barriers through the Whittemore and Knafl [61] 5-stage methodology. However, to be applied to future research, a more comprehensive perspective and a detailed guide were needed. Thus, our study not only conducted data extraction and analysis from the selected studies but also presented an integrated framework, the IAF, that can be applied to future works.

In step 2, the factors and dimensions were defined. Existing studies have proposed various frames for analyzing the acceptance factors of technologies. For instance, Schulz et al [62] presented three aspects for understanding aging and technology as follows: (1) user characteristics, (2) technology functions, and (3) social factors. Our study presents five dimensions to analyze the acceptance factors for eHealth: *personal*, *user–technology relational*, *technological*, *service-related*, and *environmental*. Compared with existing research, our framework considers not only the main dimensions (user, technology, and environment) but also the issues that emerge as a result of the interaction between the main dimensions. As can be seen in the results of the analysis (step 3; Figure 6), the acceptance factors in the *user–technology relational* dimension were most significantly addressed by the reference studies, and this dimension is estimated to be a category that should be treated as important in future studies (step 4; Figures 11 and 12). Similarly, the *service-related*

dimension reflects the fact that eHealth technologies are not isolated but integrated into health services that will also be transformed as a result of the adoption of the technology. This dimension makes visible the relevant interaction between the technology and the environment and the impact of this interaction on the user. The proposed 23 acceptance factors and 5 dimensions emerged from the existing evidence by thematic analysis in step 2, forming a framework. This IAF contributes to the state of the art a finer-grained analysis tool that integrates the diversity of elements considered in previous research and provides a common vocabulary.

In step 3, a new metric for evaluating the quality of the existing evidence was presented to expand the IAF with information about the impact of the different factors. Various methods for evaluating the quality of evidence in systematic reviews for health care have been studied [17,18]. A high score obtained through this evaluation method means that there are many confirmations that the actual effect is similar to the effect estimated by the study, whereas a low score means that the actual effect may be significantly different from the estimated effect [17]. In the systematic review on health information technology adoption [18], a ranking of 5 steps was applied to evaluate the quality of evidence, considering the study design and research method.

Although the research method and design are of paramount importance for judging the quality of evidence, additional criteria were needed that were more aligned with our analysis objective. Thus, three primary criteria were suggested by this study: the reliability of the research method, the degree of understanding of the target technology, and the year of the study. These 3 indicators contain detailed evaluation criteria that can be applied to other systematic studies as well. It is a reproducible and transparent framework for evaluating the certainty of the evidence, which minimizes author bias. Through this evaluation, the impact of acceptance factors was organized and analyzed. We believe that this is a more meaningful analysis than the assessment of the impact of the acceptance factors by just the frequency with which they were considered in previous studies. Once the impact of the acceptance factors has been incorporated into the IAF, it becomes evident that not all 23 factors are equally important for the acceptance of eHealth technologies by older users. A classical Pareto analysis was performed to select the set of highest priority factors (those that together account for 80% of the total impact). This is valuable information when facing a concrete situation seeking the adoption of a specific eHealth technology. This would help decision-makers to focus on available resources on the most influential factors.

Adaptation of the IAF

In step 4, the diversity of adoption situations considered in the previous research was addressed. The goal was to propose a way in which IAF can be adapted to reflect existing evidence in specific adoption situations and investigate the extent to which IAF is a robust analysis framework. The differences in the IAF according to the health status of participants and the verification time at which acceptance factors were identified (before or after adoption) were compared through stratification analysis. We concluded that the highest priority acceptance factors vary according to the target's health status. In addition, the data from this analysis illustrate the key acceptance factors in the IAF according to the health status of the users. Further diversity in user profiles could not be analyzed in the existing evidence. It would be interesting to explore in future research the way in which the weights of factors and dimensions differ for the same product or service based on the profile of the users.

The differences in the IAF according to the verification time were examined as well. In a previous study on the acceptance of electronic technology for older users [63], the acceptance factors of the preimplementation and postimplementation stages were compared and analyzed. Some of the factors were considered regardless of the stage; however, the acceptance factors that differed for each stage were also identified. In the same manner, in this study, we confirmed that there is a difference between the acceptance factors found relevant in the preadoption and postadoption phases of eHealth. This result confirms that it is necessary to apply the differentiated acceptance factors to different verification times.

We were also concerned about the validity of aggregating evidence about acceptance factors coming from studies performed in different years. Recently, eHealth technology has been further developed, and investment per country has increased significantly [3]. In addition, as the use rates of smartphones and the internet in older adults were different 10 years ago than they are in the present, we assumed that there would be a difference in acceptance factors with the passage of each year. As a result of comparing the acceptance factors by year in our last analysis, the continuous increase of the acceptance factors in the *user–technology relational* and the *service-related* dimensions was confirmed, except in 2018, a year in which there were few selected studies. These results support the inclusion of both dimensions in the IAF. In particular, the factors of *experience with technology* and *attitude toward technology* in the *user–technology relational* dimension have recently been dealt with as important, and the *support for use* factor in the *service-related* dimension is also a factor that has become more important in recent years. This result implies that one of the things a service provider needs to care about is ensuring that appropriate technical training is provided so that older users can use this equipment as efficiently as possible [62]. Although there is a belief that older people are not interested in using technology, many studies reveal that the facts prove otherwise and, more importantly, that there is a barrier to use because of a lack of adequate training and technical support [9]. When support, human factors, and stable technology are well combined, the barriers can be overcome. The results of this analysis identify the acceptance factors that have been becoming more important and suggest trends that can guide to set the direction for future studies.

Limitations

To ensure the validity of this study, three threats were considered: (1) selection bias in systematic reviews, (2) threats to the extraction of acceptance factors, and (3) limitations in assessing the evidence quality.

Selection Bias

There is a risk that individual bias will be reflected in the research selection. To minimize this, the 4 exclusion criteria and 5 evaluation questions were defined. On the basis of these criteria, significant effort was put into the process of selecting high-quality studies that are suitable for the subject, following the PRISMA process. During the study selection process, Covidence was used to thoroughly verify that the selection was not based on the individual opinions of the 3 authors.

Threats to the Extraction of Acceptance Factors

There is a potential threat that the terms used to explain the acceptance factors may have been used to convey different meanings in different studies. For the initial 105 factors extracted, the authors made an effort to identify the exact meaning of the terms used within the context of each study. The factors that could cause semantic confusion were analyzed through a second review and discussion of the original studies by the authors. In addition, the conjunction of expertise of the authors in different disciplines minimized a possible bias related to a limited view of the technology and themes extracted.

Limitations in Assessing the Quality of Evidence

There is a potential risk of bias in the assessment of the quality of evidence in each study. As a form of prevention, the evaluation criteria used in the previous studies were thoroughly analyzed, and the evaluation criteria proposed for this review were established to be as specific as possible. In addition, to increase the confidence that the same results can be obtained even after re-evaluation, there was a focus on objectifying the criteria.

Conclusions

Plans for and investments in eHealth are expanding worldwide. This is considered a good solution for covering places where traditional health care services do not reach [3]. The application and development of eHealth are growing more important because of the continuous increase in the older population and the availability of solutions. The continuous monitoring of older adults can reduce sudden accidents and help respond immediately to emergencies [64]. Despite their needs, the use of eHealth technology is a new challenge for older people. To pervade in daily life, not only must the technical, service-related, and environmental infrastructures of the eHealth service be prepared but also the personal factors and user–technology relational factors be considered. This is the key to understanding and addressing the needs and characteristics of users more clearly. This study systematically reviewed the research that evaluated the acceptance factors for older people in the eHealth service field over the last 11 years. In addition, this study proposed an IAF through thematic analysis and the assessment of the impact of these factors. In addition, the eHealth acceptance factors were compared and analyzed according to the participants' health conditions, verification time, and year. We expect that the IAF will become a tool that can be used to predict the main barriers to be overcome and facilitators to be leveraged. These data will form a good research material base for the application of eHealth to older users in the future.

Acknowledgments

This research was partially funded by the Spanish Ministry of Science and Innovation research grant PID2019-108408RB-C21—Active aging with Unobtrusive Personalized monitoring project and the European Project POSITIVE (Maintaining and Improving the Intrinsic Capacity Involving Primary Care and Caregivers; reference 20683) funded by the European Institute of Innovation & Technology Health.

The authors would like to thank the Fondo Europeo de Desarrollo Regional (European Regional Development Fund) funds for cofinancing our home institution.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Study overview and methods.
[DOCX File , 29 KB - jmir_v24i1e31920_app1.docx]

Multimedia Appendix 2
General distribution of the studies.
[PNG File , 94 KB - jmir_v24i1e31920_app2.png]

References

1. Ageing and health. World Health Organization. 2018. URL: https://www.who.int/en/news-room/fact-sheets/detail/ageing-and-health [accessed 2021-09-21]
2. COVID-19 risks and vaccine information for older adults. Centers for Disease Control and Prevention. 2021. URL: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html [accessed 2021-09-21]

3. Global diffusion of eHealth: making universal health coverage achievable: report of the third global survey on eHealth. World Health Organization. 2016. URL: https://www.who.int/publications/i/item/9789241511780 [accessed 2022-01-25]
4. Eysenbach G. What is e-health? J Med Internet Res 2001 Jun;3(2):E20 [FREE Full text] [doi: 10.2196/jmir.3.2.e20] [Medline: 11720962]
5. Farahani B, Firouzi F, Chang V, Badaroglu M, Constant N, Mankodiya K. Towards fog-driven IoT eHealth: promises and challenges of IoT in medicine and healthcare. Future Gener Comput Syst 2018 Jan;78:659-676. [doi: 10.1016/j.future.2017.04.036]
6. Nazir S, Ali Y, Ullah N, García-Magariño I. Internet of things for healthcare using effects of mobile computing: a systematic literature review. Wirel Commun Mob Comput 2019 Nov 14;2019:1-20. [doi: 10.1155/2019/5931315]
7. Kańtoch E, Kańtoch A. What features and functions are desired in telemedical services targeted at polish older adults delivered by wearable medical devices?-Pre-COVID-19 flashback. Sensors (Basel) 2020 Sep 11;20(18):5181 [FREE Full text] [doi: 10.3390/s20185181] [Medline: 32932848]
8. Xie B. Effects of an eHealth literacy intervention for older adults. J Med Internet Res 2011 Nov;13(4):e90 [FREE Full text] [doi: 10.2196/jmir.1880] [Medline: 22052161]
9. Charness N, Boissy P, Demiris G, Krupinski EA, Lai AM, Lopez AM. How human factors can influence the elderly in the use of telemedicine. Telemed J E Health 2010 Oct;16(8):860-866. [doi: 10.1089/tmj.2010.9948] [Medline: 20958209]
10. Young R, Willis E, Cameron G, Geana M. "Willing but unwilling": attitudinal barriers to adoption of home-based health information technology among older adults. Health Informatics J 2014 Jun;20(2):127-135. [doi: 10.1177/1460458213486906] [Medline: 24056750]
11. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989 Sep;13(3):319. [doi: 10.2307/249008]
12. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425. [doi: 10.2307/30036540]
13. Alsswey A, Al-Samarraie H. Elderly users' acceptance of mHealth user interface (UI) design-based culture: the moderator role of age. J Multimodal User Interfaces 2019 Jul 20;14(1):49-59. [doi: 10.1007/s12193-019-00307-w]
14. van Houwelingen CT, Ettema RG, Antonietti MG, Kort HS. Understanding older people's readiness for receiving telehealth: mixed-method study. J Med Internet Res 2018 Apr 06;20(4):e123 [FREE Full text] [doi: 10.2196/jmir.8407] [Medline: 29625950]
15. Zhou M, Zhao L, Kong N, Campy KS, Qu S, Wang S. Factors influencing behavior intentions to telehealth by Chinese elderly: an extended TAM model. Int J Med Inform 2019 Jun;126:118-127. [doi: 10.1016/j.ijmedinf.2019.04.001] [Medline: 31029253]
16. Cimperman M, Makovec BM, Trkman P. Analyzing older users' home telehealth services acceptance behavior-applying an Extended UTAUT model. Int J Med Inform 2016 Jun;90:22-31. [doi: 10.1016/j.ijmedinf.2016.03.002] [Medline: 27103194]
17. Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011 Apr;64(4):383-394. [doi: 10.1016/j.jclinepi.2010.04.026] [Medline: 21195583]
18. Kavandi H, Jaana M. Factors that affect health information technology adoption by seniors: a systematic review. Health Soc Care Community 2020 May 07:1827-1842. [doi: 10.1111/hsc.13011] [Medline: 32378769]
19. Tague N. The Quality Toolbox. 2nd Ed. Milwaukee, Wisconsin: ASQ Quality Press; 2005.
20. Jaana M, Sherrard H, Paré G. A prospective evaluation of telemonitoring use by seniors with chronic heart failure: adoption, self-care, and empowerment. Health Informatics J 2019 Dec;25(4):1800-1814 [FREE Full text] [doi: 10.1177/1460458218799458] [Medline: 30247080]
21. Mercer K, Giangregorio L, Schneider E, Chilana P, Li M, Grindrod K. Acceptance of commercially available wearable activity trackers among adults aged over 50 and with chronic illness: a mixed-methods evaluation. JMIR Mhealth Uhealth 2016;4(1):e7 [FREE Full text] [doi: 10.2196/mhealth.4225] [Medline: 26818775]
22. Lang C, Voigt K, Neumann R, Bergmann A, Holthoff-Detto V. Adherence and acceptance of a home-based telemonitoring application used by multi-morbid patients aged 65 years and older. J Telemed Telecare 2020 Feb 02:1357633X20901400 [FREE Full text] [doi: 10.1177/1357633X20901400] [Medline: 32009577]
23. Zettel-Watson L, Tsukerman D. Adoption of online health management tools among healthy older adults: an exploratory study. Health Informatics J 2016 Jun;22(2):171-183. [doi: 10.1177/1460458214544047] [Medline: 25149210]
24. Currie M, Philip LJ, Roberts A. Attitudes towards the use and acceptance of eHealth technologies: a case study of older adults living with chronic pain and implications for rural healthcare. BMC Health Serv Res 2015 Apr;15:162 [FREE Full text] [doi: 10.1186/s12913-015-0825-0] [Medline: 25888988]
25. Pywell J, Vijaykumar S, Dodd A, Coventry L. Barriers to older adults' uptake of mobile-based mental health interventions. Digit Health 2020;6:2055207620905422 [FREE Full text] [doi: 10.1177/2055207620905422] [Medline: 32110429]
26. Kim S, Choudhury A. Comparison of older and younger adults' attitudes toward the adoption and use of activity trackers. JMIR Mhealth Uhealth 2020 Oct 22;8(10):e18312 [FREE Full text] [doi: 10.2196/18312] [Medline: 33090116]

27. De Veer AJ, Peeters JM, Brabers AE, Schellevis FG, Rademakers JJ, Francke AL. Determinants of the intention to use e-Health by community dwelling older people. BMC Health Serv Res 2015 Mar 15;15:103 [FREE Full text] [doi: 10.1186/s12913-015-0765-8] [Medline: 25889884]
28. Herrmann M, Boehme P, Hansen A, Jansson K, Rebacz P, Ehlers JP, et al. Digital competencies and attitudes toward digital adherence solutions among elderly patients treated with novel anticoagulants: qualitative study. J Med Internet Res 2020 Jan 24;22(1):e13077 [FREE Full text] [doi: 10.2196/13077] [Medline: 32012049]
29. Cao Y, Li J, Qin X, Hu B. Examining the effect of overload on the mHealth application resistance behavior of elderly users: an SOR perspective. Int J Environ Res Public Health 2020 Sep 12;17(18) [FREE Full text] [doi: 10.3390/ijerph17186658] [Medline: 32932679]
30. Vaziri DD, Giannouli E, Frisiello A, Kaartinen N, Wieching R, Schreiber D, et al. Exploring influencing factors of technology use for active and healthy ageing support in older adults. Behav Inform Technol 2019 Jul 22;39(9):1011-1021. [doi: 10.1080/0144929x.2019.1637457]
31. Russell TG, Gillespie N, Hartley N, Theodoros D, Hill A, Gray L. Exploring the predictors of home telehealth uptake by elderly Australian healthcare consumers. J Telemed Telecare 2015 Dec;21(8):485-489. [doi: 10.1177/1357633X15606264] [Medline: 26391512]
32. Cajita MI, Hodgson NA, Lam KW, Yoo S, Han H. Facilitators of and barriers to mHealth adoption in older adults with heart failure. Comput Inform Nurs 2018 Aug;36(8):376-382. [doi: 10.1097/CIN.0000000000000442] [Medline: 29742549]
33. Li L, Peng W, Kononova A, Bowen M, Cotten SR. Factors associated with older adults' long-term use of wearable activity trackers. Telemed J E Health 2020 Jun;26(6):769-775. [doi: 10.1089/tmj.2019.0052] [Medline: 31553281]
34. Peeters JM, De Veer AJ, Van der Hoek L, Francke AL. Factors influencing the adoption of home telecare by elderly or chronically ill people: a national survey. J Clin Nurs 2012 Nov;21(21-22):3183-3193. [doi: 10.1111/j.1365-2702.2012.04173.x] [Medline: 22827253]
35. Wade R, Cartwright C, Shaw K. Factors relating to home telehealth acceptance and usage compliance. Risk Manag Healthc Policy 2012;5:25-33 [FREE Full text] [doi: 10.2147/RMHP.S30204] [Medline: 22570580]
36. Li J, Ma Q, Chan AH, Man S. Health monitoring through wearable technologies for older adults: smart wearables acceptance model. Appl Ergon 2019 Feb;75:162-169. [doi: 10.1016/j.apergo.2018.10.006] [Medline: 30509522]
37. Askari M, Klaver NS, van Gestel TJ, van de Klundert J. Intention to use medical apps among older adults in the Netherlands: cross-sectional study. J Med Internet Res 2020 Sep 04;22(9):e18080 [FREE Full text] [doi: 10.2196/18080] [Medline: 32624465]
38. Cajita MI, Hodgson NA, Budhathoki C, Han H. Intention to use mHealth in older adults with heart failure. J Cardiovasc Nurs 2017 Feb 28:1-7. [doi: 10.1097/JCN.0000000000000401] [Medline: 28248747]
39. Quaosar GM, Hoque MR, Bao Y. Investigating factors affecting elderly's intention to use m-Health services: an empirical study. Telemed J E Health 2018 Apr;24(4):309-314. [doi: 10.1089/tmj.2017.0111] [Medline: 28976824]
40. Lan Z, Liu H, Yang C, Liu X, Sorwar G. Investigating influencing factors of chinese elderly users? Intention to adopt mHealth based on the UTAUT2 model. In: Proceedings of the Fourth International Conference on Biological Information and Biomedical Engineering. 2020 Presented at: BIBE2020: The Fourth International Conference on Biological Information and Biomedical Engineering; July 21 - 23, 2020; Chengdu China p. 1-5. [doi: 10.1145/3403782.3403798]
41. Meng F, Guo X, Peng Z, Lai K, Zhao X. Investigating the adoption of mobile health services by elderly users: trust transfer model and survey study. JMIR Mhealth Uhealth 2019 Jan 08;7(1):e12269. [doi: 10.2196/12269] [Medline: 30622092]
42. Parker SJ, Jessel S, Richardson JE, Reid MC. Older adults are mobile too! Identifying the barriers and facilitators to older adults' use of mHealth for pain management. BMC Geriatr 2013;13:43 [FREE Full text] [doi: 10.1186/1471-2318-13-43] [Medline: 23647949]
43. Cimperman M, Brenčič MM, Trkman P, Stanonik MD. Older adults' perceptions of home telehealth services. Telemed J E Health 2013 Oct;19(10):786-790 [FREE Full text] [doi: 10.1089/tmj.2012.0272] [Medline: 23931702]
44. Axelsson SW, Wikman AM. Ready for eHealth. Older Swedes' prceptions of eHealth services: using the PIADS Scale as a predictor for readiness. Technologies 2016 Sep 07;4(3):29. [doi: 10.3390/technologies4030029]
45. Lin TT, Bautista JR, Core R. Seniors and mobiles: a qualitative inquiry of mHealth adoption among Singapore seniors. Inform Health Soc Care 2020 Oct 01;45(4):360-373. [doi: 10.1080/17538157.2020.1755974] [Medline: 32484720]
46. Charness N, Best R, Evans J. Supportive home health care technology for older adults: attitudes and implementation. Gerontechnology 2016 Feb 23;15(4):233-242 [FREE Full text] [doi: 10.4017/gt.2016.15.4.006.00] [Medline: 29033700]
47. Tsai T, Lin W, Chang Y, Chang P, Lee M. Technology anxiety and resistance to change behavioral study of a wearable cardiac warming system using an extended TAM for older adults. PLoS One 2020;15(1):e0227270 [FREE Full text] [doi: 10.1371/journal.pone.0227270] [Medline: 31929560]
48. Rockmann R, Gewald H. Technology-mediated health activities: an exploratory study on older adults. In: Proceedings of the 22nd Americas Conference on Information Systems (AMCIS). 2016 Presented at: 2nd Americas Conference on Information Systems (AMCIS); August 11-14, 2016; San Diego, CA, USA p. 1273-1282 URL: https://publications.hs-neu-ulm.de/333/
49. Buccoliero L, Bellio E. The adoption of "silver" e-Health technologies: first hints on technology acceptance factors for elderly in Italy. In: Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance.:

Association for Computing Machinery; 2014 Presented at: ICEGOV2014: 8th International Conference on Theory and Practice of Electronic Governance; October 27 - 30, 2014; Guimaraes Portugal p. 304-307. [doi: 10.1145/2691195.2691303]
50. Best R, Souders DJ, Charness N, Mitzner TL, Rogers WA. The role of health status in older adults' perceptions of the usefulness of eHealth technology. Hum Asp IT Aged Popul (2015) 2015 Sep;9194:3-14 [FREE Full text] [doi: 10.1007/978-3-319-20913-5_1] [Medline: 31240277]
51. Knapova L, Klocek A, Elavsky S. The role of psychological factors in older adults' readiness to use eHealth technology: cross-sectional questionnaire study. J Med Internet Res 2020 May 28;22(5):e14670 [FREE Full text] [doi: 10.2196/14670] [Medline: 32348251]
52. Dolnicar V, Petrovcic A, Setinc M, Kosir I, Kavcic M. Understanding acceptance factors for using e-care systemsdevices: insights from a mixed-method intervention study in Slovenia. In: Human Aspects of IT for the Aged Population. Cham: Springer; 2017:362-377.
53. Hoque R, Sorwar G. Understanding factors influencing the adoption of mHealth by the elderly: an extension of the UTAUT model. Int J Med Inform 2017 May;101:75-84. [doi: 10.1016/j.ijmedinf.2017.02.002] [Medline: 28347450]
54. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User acceptance of wrist-worn activity trackers among community-dwelling older adults: mixed method study. JMIR Mhealth Uhealth 2017 Nov 15;5(11):e173 [FREE Full text] [doi: 10.2196/mhealth.8211] [Medline: 29141837]
55. Ware P, Bartlett SJ, Paré G, Symeonidis I, Tannenbaum C, Bartlett G, et al. Using eHealth technologies: interests, preferences, and concerns of older adults. Interact J Med Res 2017 Mar 23;6(1):e3 [FREE Full text] [doi: 10.2196/ijmr.4447] [Medline: 28336506]
56. Diño MJ, de Guzman AB. Using Partial Least Squares (PLS) in predicting behavioral intention for telehealth use among Filipino elderly. Edu Gerontol 2014 Jun 25;41(1):53-68. [doi: 10.1080/03601277.2014.917236]
57. Portz JD, Bayliss EA, Bull S, Boxer RS, Bekelman DB, Gleason K, et al. Using the technology acceptance model to explore user experience, intent to use, and use behavior of a patient portal among older adults with multiple chronic conditions: descriptive qualitative study. J Med Internet Res 2019 Apr 08;21(4):e11604 [FREE Full text] [doi: 10.2196/11604] [Medline: 30958272]
58. Nymberg VM, Bolmsjö BB, Wolff M, Calling S, Gerward S, Sandberg M. 'Having to learn this so late in our lives…' Swedish elderly patients' beliefs, experiences, attitudes and expectations of e-health in primary health care. Scand J Prim Health Care 2019 Mar;37(1):41-52 [FREE Full text] [doi: 10.1080/02813432.2019.1570612] [Medline: 30732519]
59. Rooney J, Kubiak T, Westcott R, Reid R, Wagoner K, Pylipow P. Building from the basics: Master these quality tolls and do your job better. Quality Engineering 2009;54:229-230.
60. Foster MV, Sethares KA. Facilitators and barriers to the adoption of telehealth in older adults: an integrative review. Comput Inform Nurs 2014 Nov;32(11):523-534. [doi: 10.1097/CIN.0000000000000105] [Medline: 25251862]
61. Whittemore R, Knafl K. The integrative review: updated methodology. J Adv Nurs 2005 Dec;52(5):546-553. [doi: 10.1111/j.1365-2648.2005.03621.x] [Medline: 16268861]
62. Schulz R, Wahl H, Matthews JT, De Vito DA, Beach SR, Czaja SJ. Advancing the aging and technology agenda in gerontology. Gerontologist 2015 Oct;55(5):724-734. [doi: 10.1093/geront/gnu071] [Medline: 25165042]
63. Peek ST, Wouters EJ, van Hoof J, Luijkx KG, Boeije HR, Vrijhoef HJ. Factors influencing acceptance of technology for aging in place: a systematic review. Int J Med Inform 2014 Apr;83(4):235-248 [FREE Full text] [doi: 10.1016/j.ijmedinf.2014.01.004] [Medline: 24529817]
64. Maswadi K, Ghani NB, Hamid SB. Systematic literature review of smart home monitoring technologies based on iot for the elderly. IEEE Access 2020;8:92244-92261. [doi: 10.1109/access.2020.2992727]

Abbreviations

GRADE: grading of recommendations, assessment, development, and evaluation
IAF: integrated acceptance framework
IoT: Internet of Things
mHealth: mobile health
PICO: population, intervention, comparison, and outcomes
POSITIVE: Maintaining and Improving the Intrinsic Capacity Involving Primary Care and Caregivers
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TAM: Technology Acceptance Model
UTAUT: Unified Theory of Acceptance and Use of Technology

Edited by A Mavragani; submitted 09.07.21; peer-reviewed by V Traver, J Offermann-van Heek; comments to author 12.08.21; revised version received 21.09.21; accepted 27.11.21; published 28.01.22.

Please cite as:
Yu J, de Antonio A, Villalba-Mora E
Design of an Integrated Acceptance Framework for Older Users and eHealth: Influential Factor Analysis
J Med Internet Res 2022;24(1):e31920
URL: https://www.jmir.org/2022/1/e31920
doi:10.2196/31920
PMID:35089155

Original Paper

Crowdsourcing for Machine Learning in Public Health Surveillance: Lessons Learned From Amazon Mechanical Turk

Zahra Shakeri Hossein Abad[1,2], PhD; Gregory P Butler[3], MSc; Wendy Thompson[3], MSc; Joon Lee[2,4,5], PhD

[1]Department of Biomedical Informatics, Harvard Medical School, Harvard University, Boston, MA, United States

[2]Data Intelligence for Health Lab, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[3]Centre for Surveillance and Applied Research, Public Health Agency of Canada, Ottawa, ON, Canada

[4]Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[5]Department of Cardiac Sciences, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

Corresponding Author:
Joon Lee, PhD
Data Intelligence for Health Lab
Cumming School of Medicine
University of Calgary
3280 Hospital Dr NW
Calgary, AB, T2N 4Z6
Canada
Phone: 1 403 220 2968
Email: joonwu.lee@ucalgary.ca

Abstract

Background: Crowdsourcing services, such as Amazon Mechanical Turk (AMT), allow researchers to use the collective intelligence of a wide range of web users for labor-intensive tasks. As the manual verification of the quality of the collected results is difficult because of the large volume of data and the quick turnaround time of the process, many questions remain to be explored regarding the reliability of these resources for developing digital public health systems.

Objective: This study aims to explore and evaluate the application of crowdsourcing, generally, and AMT, specifically, for developing digital public health surveillance systems.

Methods: We collected 296,166 crowd-generated labels for 98,722 tweets, labeled by 610 AMT workers, to develop machine learning (ML) models for detecting behaviors related to physical activity, sedentary behavior, and sleep quality among Twitter users. To infer the ground truth labels and explore the quality of these labels, we studied 4 statistical consensus methods that are agnostic of task features and only focus on worker labeling behavior. Moreover, to model the meta-information associated with each labeling task and leverage the potential of context-sensitive data in the truth inference process, we developed 7 ML models, including traditional classifiers (offline and active), a deep learning–based classification model, and a hybrid convolutional neural network model.

Results: Although most crowdsourcing-based studies in public health have often equated majority vote with quality, the results of our study using a truth set of 9000 manually labeled tweets showed that consensus-based inference models mask underlying uncertainty in data and overlook the importance of task meta-information. Our evaluations across 3 physical activity, sedentary behavior, and sleep quality data sets showed that truth inference is a context-sensitive process, and none of the methods studied in this paper were consistently superior to others in predicting the truth label. We also found that the performance of the ML models trained on crowd-labeled data was sensitive to the quality of these labels, and poor-quality labels led to incorrect assessment of these models. Finally, we have provided a set of practical recommendations to improve the quality and reliability of crowdsourced data.

Conclusions: Our findings indicate the importance of the quality of crowd-generated labels in developing ML models designed for decision-making purposes, such as public health surveillance decisions. A combination of inference models outlined and analyzed in this study could be used to quantitatively measure and improve the quality of crowd-generated labels for training ML models.

(J Med Internet Res 2022;24(1):e28749) doi:10.2196/28749

KEYWORDS

crowdsourcing; machine learning; digital public health surveillance; public health database; social media analysis

Introduction

Background

In recent years, social media data have been extensively used in different areas of public health [1-3], such as detecting outbreaks and emerging diseases [4,5], monitoring adverse drug reactions [6], and predicting or modeling health-related behaviors and outcomes [7-9]. Since 2011, Twitter has been the most popular form of social media used for public health communication [10,11]. In 2020, Twitter alone reported 500 million tweets generated per day from 145 million daily active users. A recent scoping review of 755 articles on digital public health surveillance shows that Twitter is the most studied of all platforms and the most used platform to study communicable diseases, behavioral risk factors, mental health, drug use, and vaccines [11]. In addition to the inherent limitations of social media data, such as lack of demographic data and biased populations, when integrated with complex data-driven models such as artificial neural networks, these publicly accessible resources can be used for population-level surveillance to complement traditional public health surveillance (eg, surveys) with faster and less costly longitudinal information.

Although linguistic annotation is crucial for developing machine learning (ML) and natural language processing (NLP) models, manual labeling of a large volume of data is a notorious problem because of its high cost and labor-intensive nature. In recent years, this problem has been tackled using crowdsourcing technologies such as Amazon Mechanical Turk (AMT) [12], Crowdflower [13], and Prolific Academic [13] to obtain relatively low-cost labeled data more quickly and easily. AMT is a software service operated by Amazon that allows users to crowdsource work, broken into microtasks called HITs (Human Intelligence Tasks), to a large number of workers who are compensated for each HIT completed. With the vast potential applications of crowdsourcing in public health [14-16], the research community has seen steady growth in the use of AMT in the past 10 years. The number of studies indexed in PubMed using the search term *Amazon Mechanical Turk* AND *public health* has increased sharply from 42 studies in 2015 to 118 studies in 2019.

However, because of the uncertain quality of AMT workers with unknown expertise, their labels are sometimes unreliable, forcing researchers and practitioners to collect information redundantly, which poses new challenges in the field. Given that in large-scale crowdsourcing tasks the same workers cannot label all the examples, measuring interannotator agreement and managing the quality of workers differ from those of a team of in-house expert workers. Despite the growing popularity of AMT for developing ML models in public health research, the reliability and validity of this service have not yet been investigated. At least several public health studies have used AMT for training data-driven ML models without external gold standard comparisons [17-21]. Ayers et al [17] used AMT to create a gold standard data set to develop predictive models to detect electronic nicotine delivery systems on social media. Yin et al [18] developed a scalable classifier to detect personal health mentions on Twitter based on a gold standard data set generated by AMT workers. The reliability of the crowd-labeled data set in this study was measured based on the agreement among workers.

Similarly, to characterize sleep quality using Twitter, McIver et al [19] used AMT for sentiment annotation of text data and used interannotator agreement to assess the reliability of workers. Reece et al [20] used AMT to build a data set and develop a prediction model to detect depression emergence and posttraumatic stress disorder in Twitter users. To control the quality of the data collected, they required the workers to have completed at least hundred tasks, with a minimum 95% approval rating. Although research has supported the efficacy of using reputation to evaluate the quality of crowdsourced data [22], the reliability of using this metric in developing ML-based digital public health systems has not yet been investigated. Thus, in this study, in addition to defining qualification requirements for AMT workers, we studied the reliability of crowd-generated training data for developing ML models in the context of public health surveillance. We used AMT to collect 296,166 labels for 98,722 unique tweets, labeled by 610 AMT workers, to develop ML models that can detect the physical activity, sedentary behavior, and sleep quality (PASS) of Twitter users.

Objectives

The primary aim of this study is to evaluate the application of AMT for training data-driven ML models by analyzing the quality of crowd-generated labels. As the quality of crowd-generated labels, regardless of the type of the task being studied, is critical to the robustness of ML models trained based on these labels, we created a gold standard data set of labels and applied several statistical and ML-based models to assess the reliability of using the crowd-labeling task from different perspectives (eg, process, design, and inference). To interpret the results of our quality assessment and explore the effect of noisy labels on the applicability of inference models in dealing with these labels, our approach involved evaluating the performance of 4 consensus methods, which do not involve task features in their truth inference, and exploring their feasibility in improving the quality of crowd-labeled data. As these methods are modeled purely as a function of worker behaviors concerning labeling tasks, they cannot leverage the value of context-sensitive information (ie, the task's meta-information) in their inference decisions. Thus, we collected additional features for our labeling data set and developed 7 ML models, including a deep learning (DL) model and a hybrid convolutional neural network (CNN) architecture to couple worker behaviors with the task's meta-information when inferring the truth label. To detect and correct noisy labels, we also developed 5 pool-based active learners to iteratively detect the most informative samples (ie, samples with more uncertainty) and remove them from the validation set. Finally, we used SHAP (Shapley Additive Explanations) [23] to explore the contribution of different features, including worker behaviors and

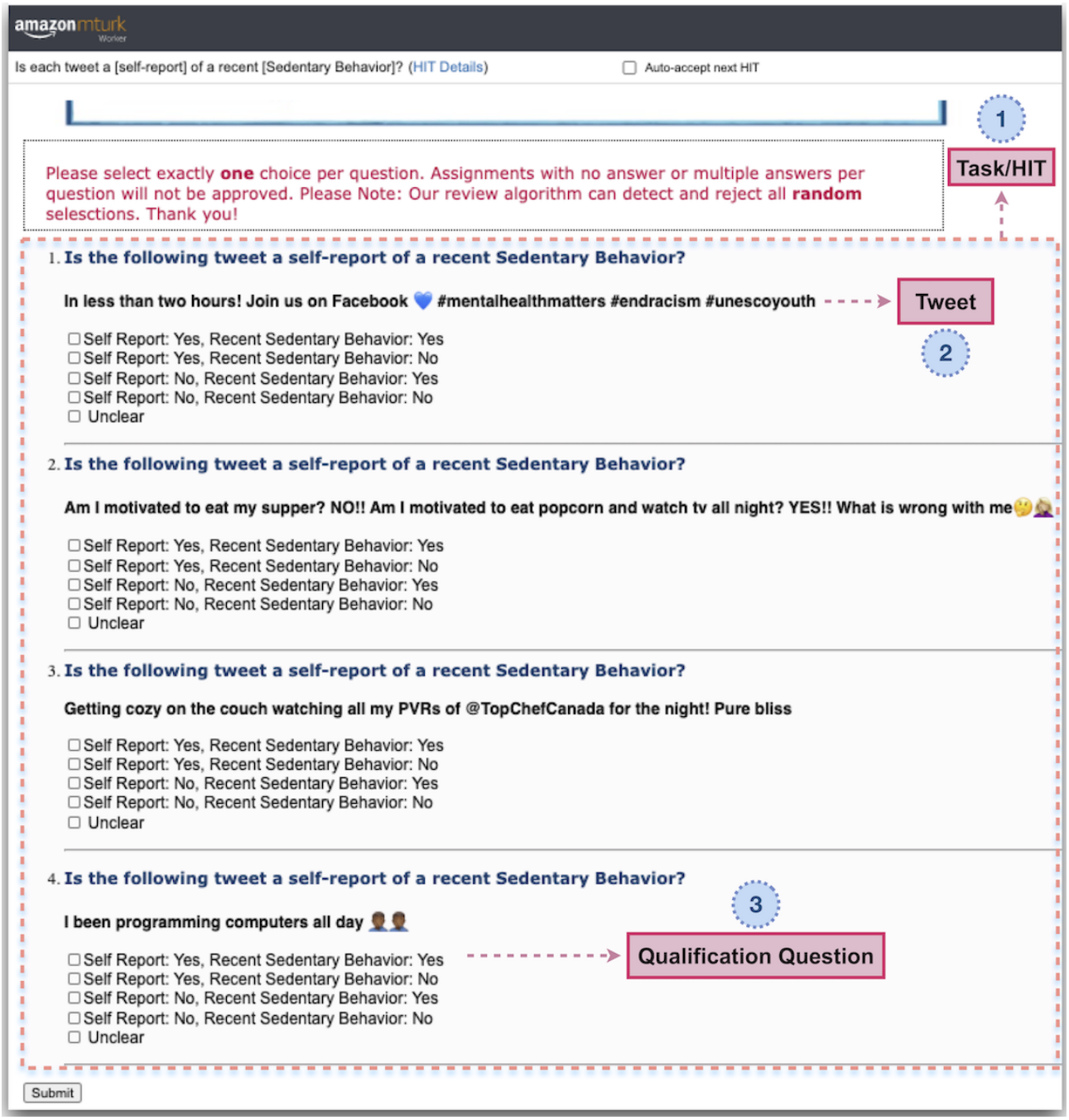
context-sensitive features, to the results of our supervised inference models.

Methods

Labels

The crowdsourcing tasks, referred to as HITs by AMT, were designed to collect 5 labels based on 2 conditions, self-reported and recent PASS experience, to develop binary and multiclass classification models that can detect PASS-related behavior in Twitter users. The labels of the multiclass prediction models were defined as 11, 10, 01, and 00, based on the value of each condition (Figure S1 in Multimedia Appendix 1). We also let workers choose a fifth option, called *unclear*, to ensure they did not give random labels to tasks they were not confident in performing successfully (Figure 1). We excluded this label for both inference and classification tasks. We defined the binary labels as 1 if both conditions were met and 0, otherwise. The binary labels did not directly come from the AMT workers and were generated by dichotomizing the collected labels.

Figure 1. A sample labeling task (ie, human intelligence task [HIT]) for sedentary behavior. Each HIT contains 4 questions (section 1), and each asks if the presented tweet is a self-reported physical activity, sedentary behavior, or sleep quality–related behavior (section 2). The fourth question is an easy, qualification question that was used to check the quality of the worker (section 3).



Crowdsourcing Workflow

We implemented a pipeline to create the HITs, post them on AMT, collect the labels through a quality check process, approve or reject the HITs, and store the results. To minimize noisy and low-quality data, we added a qualification requirement to our tasks and granted labeling access to workers who had demonstrated a high degree of success in performing a wide range of HITs across AMT (ie, master qualification). In addition, we added a simple qualification question to each HIT to detect spammers or irresponsible workers. Each HIT contained 4 questions, including the qualification question, and was assigned to 3 workers (Figure 1 and Figures S2 and S3 in Multimedia Appendix 1). Workers were asked to select exactly 1 choice per tweet, and HITs with zero or more than one label were rejected during the approval process. Through different iterations of data labeling, workers were paid from US $0.03 to US $0.05 after completing each HIT. We collected the labels for the 98,722 tweets used in this study through different iterations, from April 2019 to June 2020. We regularly checked the quality of the submitted tasks to detect low-quality workers during each iteration and revoke their access to our tasks. Before the formal initiation of the process, we pilot-tested the design, response time, and complexity of the HITs through 2 different iterations and revised the workflow accordingly. We did not collect any personally identifiable information from the workers (participants) during the data labeling task. The experiments were carried out in accordance with the relevant guidelines and the University of Calgary Conjoint Faculties Research Ethics Board regulations. We implemented the entire workflow in Python and used Boto3 Python software development kit to connect to and work with AMT.

Data Collection

We collected data for this study from Twitter using the Twitter livestream application programming interface (API) for the period between November 28, 2018, and June 30, 2020. The data set was filtered to include only Canadian tweets relevant to PASS. A total of 103,911 tweets were selected from 22,729,110 Canadian tweets using keywords and regular expressions related to PASS categories. Each of these 103,911 tweets was labeled by 3 AMT workers, from which 98,722 tweets received 3 valid labels, with almost half of them related to physical activity.

The demographic variables of age and gender and the information about the source of each tweet (eg, organization vs real users) were not available within the data set collected from Twitter. We estimated these variables for each tweet using the M3 inference package in Python [24], which uses a multimodal deep neural architecture for the joint classification of age, gender, and information sources of social media data. The text (tweet) field and each of the daytime, weekday, and month variables were extracted from the metadata provided by the Twitter API.

We have made the Twitter data set used in this study publicly available [25].

Data Processing

Tweets have a bounding box of coordinates, which enables spatial mapping to their respective city locations. As the Twitter API returns datetime values in Coordinated Universal Time, we used a time zone finder in Python and adjusted the time of each tweet based on its spatial data. Given that daytime, month, and weekday can be influential factors in twitting about each of the PASS categories, and to better use the datetime data (%a %b %d %H: %M: %S %Y), we extracted a: weekday, b: month, and H: hour fields and stored them as separate features.

We cleaned the text column by eliminating all special characters (eg, #, &, and @), punctuations, weblinks, and numbers. We also replaced common contractions with their uncontracted forms; for example, *I'll* was resolved as *I will*. While developing and evaluating our NLP models, we noticed that the impact of removing stop words, stemming, and converting the text to lower case on the performance of our predictive models was not noticeable. This could relate to the ability of transfer-learning techniques (ie, GloVe embeddings) to generalize on unseen data. Thus, we applied neither stop-word removing nor lexical cleaning on the textual features of our data set. Moreover, as hashtags and emojis can be used as independent words and facilitate emotional expressions, we did not remove them during the cleaning process.

To develop the ML models, all categorical data were encoded into dummy variables using one-hot encoding, and as we only approved HITs with complete answers, this data set did not contain any missing data.

Label Consistency

To measure the consistency of answers given by the workers, we calculated label consistency (LC) as the average entropy of the collected labels for each PASS category [26]. For each tweet $t_i \in T_s$, where T_s denotes the set of all tweets related to surveillance category $s \in$ {physical activity, sleep quality, sedentary behavior}, n_{ij} defines the number of answers given to the j^{th} choice ($j \in$ {1,2,3,4,5}, as we had 5 choices for each tweet). We calculated LC_s as follows:

|s| denotes the size of the surveillance category *s* and, as we collected 3 labels for each tweet, the denominators in the entropy formula received a constant value of 3. *LC* ranges from 0 to 1, and values close to 1 show more consistency among the workers' input.

Ground Truth Data Set

To investigate the viability of unsupervised inference models in predicting truth labels from crowd-labeled data and compare it with that of supervised predictive models, we used a random sample of our data set as a ground truth set (ie, 9000 tweets: 4000 tweets for physical activity, 3000 tweets for sleep quality, and 2000 tweets for sedentary behavior). In total, 6 data scientists manually labeled this sample, and the entire labeled data set was reviewed manually and relabeled by an experienced in-house domain expert in both ML and public health surveillance. The disagreements between this data set and the

crowd-labeled data set were manually checked to exclude any labeling bias that could impact the results of this study.

Inference Models

The majority voting (MV) approach estimates the actual ground truth based on most labels submitted by different workers. For example, defining the estimated label as , and the submitted label by worker w as l_w, the MV approach, for a binary labeling task, assigns 1 to if and 0, otherwise. Although individual workers' reliability coming from different backgrounds with different quality levels varies, the MV approach assumes equal expertise among the workers and does not model worker behaviors [27]. As this approach is completely task-independent, it does not involve task properties in the inference process; thus, it is fast.

The David and Skene (DS) [28] approach uses expectation–maximization (EM) to simultaneously estimate the error rate of annotators (workers) and latent label classes, when, similar to MV, the ground truth is unknown, and workers are assumed to operate independently. Unlike MV, which is agnostic to worker behavior, DS models worker k's behavior as a function of each task's true label by creating a confusion matrix π^k with size $L \times L$, where L is a fixed number and represents the number of possible labels for a single-labeled classification task. DS defines worker k's error rate as follows:

As not all workers need to label all the tasks, and a worker may label the same task more than once, sparsity can be a problem in large-scale labeling tasks when using the DS approach [27]. DS iteratively estimates the true label of each task based on the worker's quality and estimates the worker's error rate (quality) based on the inferred labels until it converges. Although the worker-specific confusion matrix generates the quality score of each worker, it may not be sufficient to measure the actual contribution of each worker [29]. The inherent complexity of a task, especially in NLP, or a worker's bias may result in wrong labels, although the worker is quantitatively accurate.

The generative model of labels, abilities, and difficulties (GLAD) [30] models the quality of workers as a function of the input task using parameter α. The quality parameter ranges from $-\infty$ to $+\infty$, implying that the worker always labels the tasks incorrectly or correctly, respectively. When $\alpha=0$, the worker cannot distinguish among the labels, and their input does not contribute to the task's correct label. To estimate the ground truth, in addition to the workers' quality, GLAD models the difficulty of task t_i as $d_i=1/\beta_i$, where $\beta_i>0$. The difficulty index ranges from 0 to ∞, where $d_i=\infty$ classifies t_i as the most difficult task, and $d_i=0$ means that the task always receives a correct label, even from the workers with $\alpha\leq 0$. GLAD uses the EM approach to obtain the maximum likelihood estimation of α and β, and models the probability that worker k correctly labels t_i using .

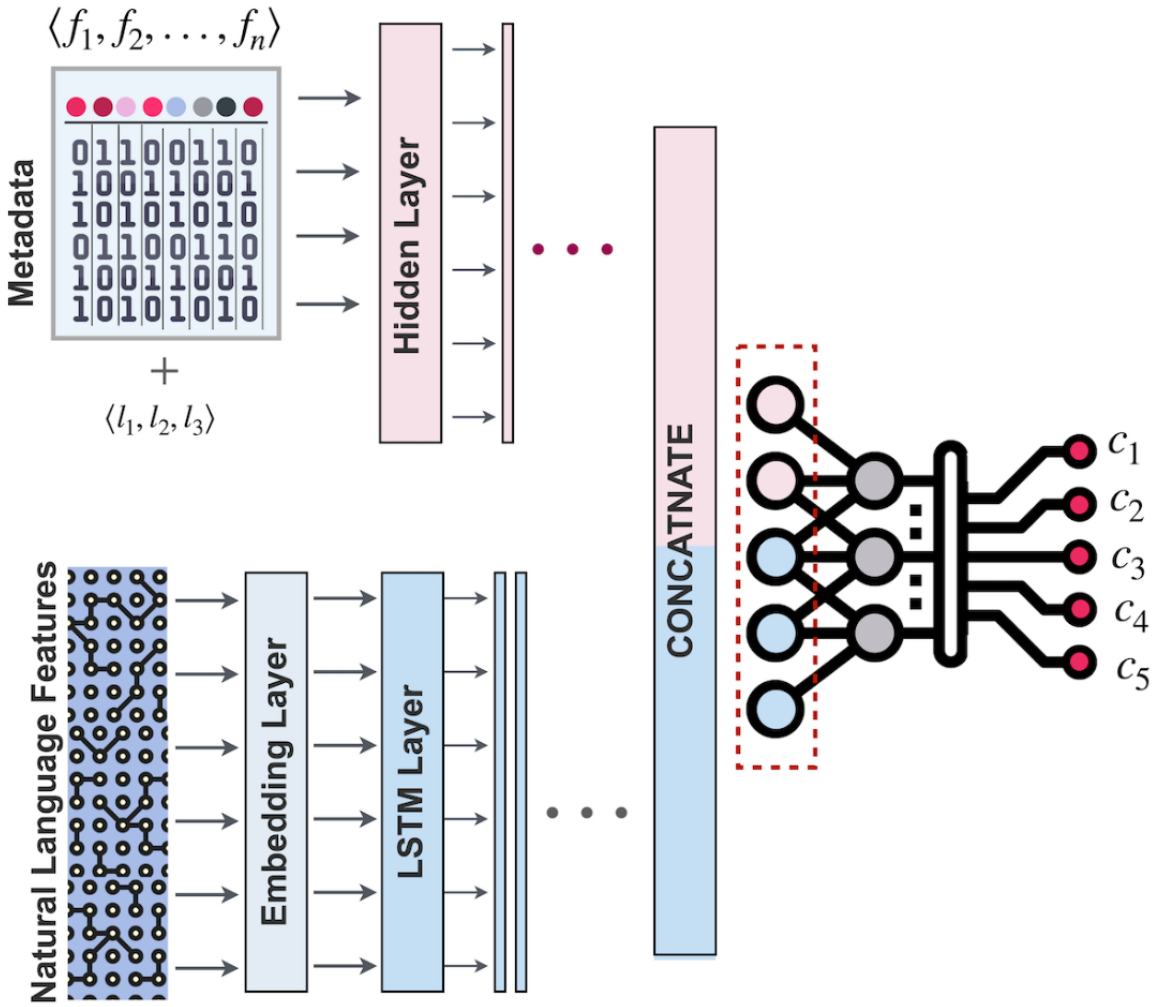
Similar to DS, Raykar algorithm (RY) [31] forms a confusion matrix to model a worker's quality. In addition, in the case of binary classification, it models worker's bias toward the positive class (ie, sensitivity) and toward the negative class (ie, specificity) using beta prior [27]. Worker bias in this context usually occurs when a worker underestimates or overestimates the truth of a task [26]. As with DS and GLAD, RY uses an unsupervised EM approach to estimate each of the model parameters and truth labels. Depending on the availability of task-specific features, RY can either use automatic supervised classifiers or fall back to unsupervised EM models to estimate the truth label.

Predictive Models

As the meta-information associated with each task may reveal its underlying complexity and thus help model worker behaviors, we developed a set of ML models to involve this metadata in the inference process. Models were trained based on quintuple F: (W,I,M,t,l), where $W = \{w_1,...,w_k\}$ represents labels collected from AMT workers, $I = \{MV,DS,RY,GLAD\}$ denotes the results of inference models, and M denotes metadata associated with each tweet including time (ie, weekday, month, and daytime), gender, age group, and the source of the tweet (ie, organization vs real people). The text of each tweet is presented by t, and l denotes the truth label.

To mitigate the risk of biased results caused by a specific learning algorithm and overcome the overfitting problem, we developed and evaluated 5 standard ML classifiers with different architectures, including generalized linear (logistic regression [LR]), kernel-based (support vector machines [SVM]), decision-tree–based (random forest and XGBoost), and sample-based (K-nearest neighbors [KNN]) classifiers. Moreover, to incorporate textual features into our analysis, we developed a hybrid DL architecture in which a CNN based on long short-term memory (LSTM) learns textual data t and a multilayer perceptron deep neural network learns metadata (W,J,M). The cleaned text, represented as an integer-encoded vector, is converted into pretrained tweet word embeddings using GloVe [32] (containing 2 billion tweets, 27 billion tokens, and 1.2 million vocabularies) in the embedding layer. The output of this layer is passed through an LSTM layer for sequence modeling, followed by 1 dropout layer to avoid overfitting and 2 dense ReLU (Rectified Linear Unit) layers. Simultaneously, the metadata of each tweet is passed through 3 fully connected layers with ReLU activation. The outputs of these networks are concatenated into a dense layer, followed by 2 fully connected dense layers, terminating at an output layer with softmax activation, cross-entropy loss, and the adam optimizer. A high-level presentation of this architecture is shown in Figure 2.

Figure 2. The pipeline of the deep learning model used to predict labels using both textual information and meta-information. LSTM: long short-term memory.



To counter the bias caused by class imbalance, for both multiclass and binary classification tasks, we used the class-weight approach to incorporate the weight of each class into the cost function by assigning higher weights to minority classes and lower weights to the majority classes. We also used the SMOTE (Synthetic Minority Oversampling Technique-Nominal Continuous) [33] approach to oversample the minority classes by creating synthetic samples based on their feature space. However, we did not notice much difference between using and not using the synthetic minority oversampling technique. Thus, our final models were trained using the class-weight approach. The hyperparameters for each method were determined using a nested 10-fold cross-validation Bayesian optimization [34].

As the main goal of both supervised and unsupervised label inference models was to minimize the number of false-negative and false-positive inferences, to evaluate the models developed in this study, we used precision, recall, F1, and precision-recall area under the curve (AUC_{PR}) metrics.

All the computations and predictive models were implemented using Python 3.7 with TensorFlow 2.0 [35], Keras [36], and Scikit-learn [37] libraries. To facilitate the replication of our study, the code repository of this study is publicly available on GitHub [38].

Results

Raw Labels From AMT Workers

In total, 610 unique workers participated in our data labeling tasks and completed 103,911 HITs, from which 5189 HITs were removed as they did not receive 3 valid answers. We approved 98,722 tasks for further analysis. Most workers (530/610, 86.9%) completed <100 HITs, of which 164 completed only 1 HIT. Among the workers who completed >5000 HITs, 1 worker completed 21,801 HITs and 3 workers completed between 5000 and 10,000 HITs (Figure 3). The calculated *LC* for each PASS category for multiclass labeling was 0.54, 0.58, and 0.55 and for binary labeling was 0.75, 0.77, and 0.74 (Table 1). This implies a high level of label inconsistency, prompting the need for further label quality analysis for the development of ML models.

Figure 3. The number of workers who completed different numbers of human intelligence tasks (HITs). Most workers completed a relatively small number of HITs.

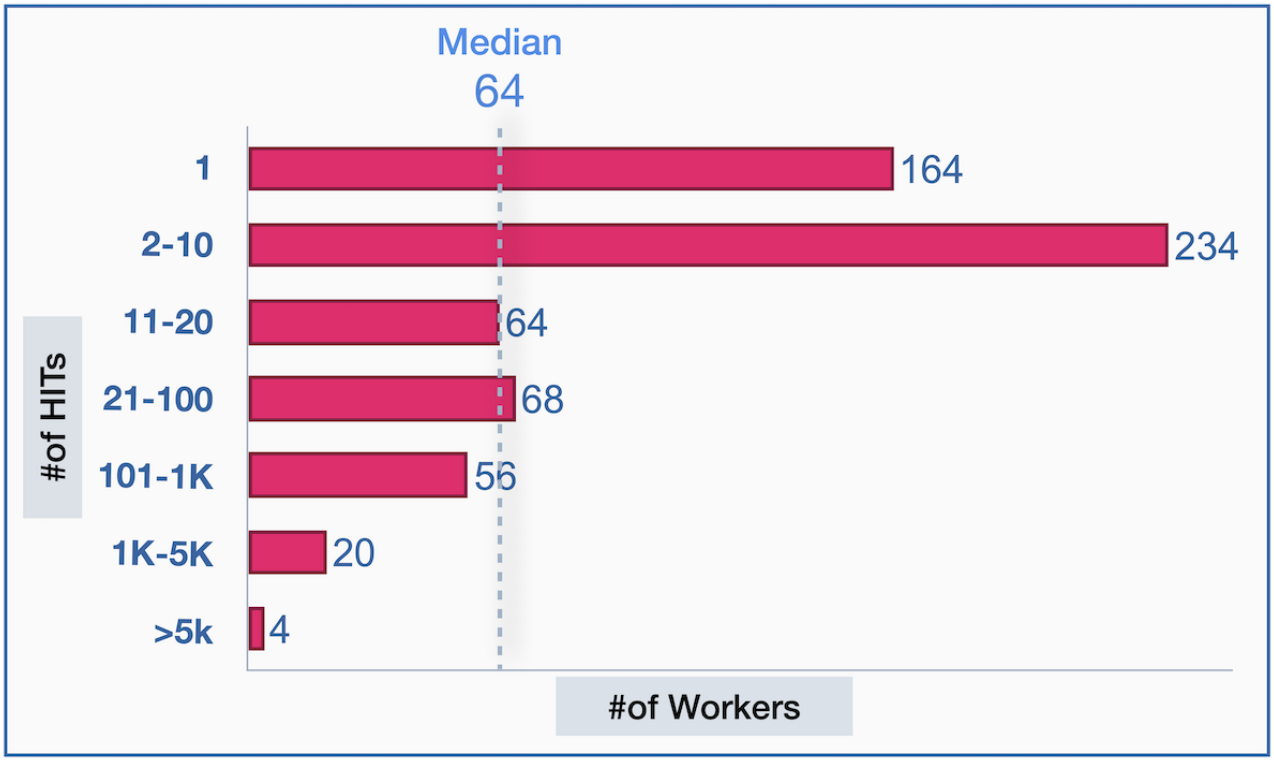


Table 1. Details of the collected labels and label consistency (LC) score for each of the physical activity, sleep quality, and sedentary behavior categories. LC ranges from 0 to 1, and the values close to 1 show more consistency among workers' input.

Type	Tweets, n (%)	LC_{multi}	LC_{binary}	Workers, n (%)
Physical activity	48,576 (49.2)	0.54	0.75	232 (38)
Sedentary behavior	17,367 (17.6)	0.55	0.74	157 (25.7)
Sleep quality	32,779 (33.2)	0.58	0.77	221 (36.2)
Total	98,722 (100)	0.56	0.75	610 (100)

Truth Inference

Table 2 describes the ground truth data set of 9000 tweets that was used to train the truth inference models. Table 3 lists the inference results obtained from the 4 unsupervised models and 7 supervised predictive models, including 2 DL models, on the ground truth data set. Each model was evaluated on both binary and multiclass versions of the data set for each PASS category. Among the unsupervised models for physical activity and sleep quality, DS and RY performed better than MV and GLAD for all performance metrics, whereas MV outperformed the other models on the sleep quality data set. Interestingly, for binary inference across all PASS categories, MV outperformed or performed just as well as the other methods, indicating the impact of task complexity on the performance of inference methods.

DL_{meta} outperforms other methods with the minimum number of false positives (precision: 78%) for the multiclass classification task, but other methods performed better with respect to recall, F1, and AUC_{PR} metrics. Performance on each PASS data set for binary classification did not highlight any individual method constantly performing best. For example, whereas SVM showed the best performance for physical activity, KNN and LR outperformed other models for sleep quality and sedentary behavior, respectively. LR achieved superior performance across all data sets for the multiclass inference task. To analyze this further, we modified the hyperparameters of the LR algorithm presented in Table 3 to stochastic average gradient solver and l_2 regularization and the optimizer of the hybrid neural network to stochastic gradient descent and repeated the comparisons. LR still outperformed the neural network model by more than 2% in all metrics. The poor performance of the neural networks in this study could be attributed to the imbalanced ratio of data (per class) to the model parameters (ie, high variance).

Table 2. Characteristics of the ground truth data set used to develop and evaluate the supervised and unsupervised inference models.

Variable	Physical activity (n=4000)	Sedentary behavior (n=2000)	Sleep quality (n=3000)
Labels, n (%)			
Binary			
Yes	1629 (40.73)	726 (36.3)	1063 (35.43)
No	2371 (59.28)	1274 (63.7)	1937 (64.57)
Multiclass			
YY[a]	1629 (40.73)	726 (36.3)	1063 (35.43)
YN[b]	550 (13.75)	395 (19.75)	862 (28.73)
NY[c]	179 (4.48)	19 (0.95)	52 (1.73)
NN[d]	1642 (41.05)	860 (43)	1023 (34.1)
Gender, n (%)			
Female	1131 (28.28)	576 (28.80)	469 (15.63)
Male	1980 (49.50)	906 (45.30)	490 (16.34)
Unknown	889 (22.22)	518 (25.90)	2041 (68.03)
Age range (years), n (%)			
≤18	204 (5.10)	170 (8.50)	150 (5)
19-29	743 (18.58)	475 (23.75)	331 (11.03)
30-39	897 (22.42)	365 (18.25)	249 (8.30)
≥40	1267 (31.68)	472 (23.60)	229 (7.64)
Unknown	889 (22.22)	518 (25.90)	2041 (68.03)
Day of week, n (%)			
Sunday	664 (16.60)	325 (16.25)	440 (14.66)
Monday	595 (14.88)	307 (15.35)	440 (14.66)
Tuesday	493 (12.32)	245 (12.25)	435 (14.50)
Wednesday	504 (12.60)	278 (13.9)	)393 (13.10)
Thursday	525 (13.12)	270 (13.50)	416 (13.86)
Friday	531 (13.28)	274 (13.70)	421 (14.03)
Saturday	668 (16.70)	283 (14.15)	2433 (14.43)
Unknown	20 (0.50)	18 (0.90)	22 (0.76)
Time (24 hours), Q1-Q3	10-19	10-19	5-18
Month (range)	February to July	April to September	January to August
Source, n (%)			
Organization	563 (14.08)	179 (8.95)	97 (3.23)
Users	3437 (85.93)	1821 (91.05)	2903 (96.77)

[a]YY: self-reported and recent physical activity, sedentary behavior, and sleep quality experience.

[b]YN: self-reported but not recent physical activity, sedentary behavior, and sleep quality experience.

[c]NY: not self-reported but recent physical activity, sedentary behavior, and sleep quality experience.

[d]NN: neither self-reported nor recent physical activity, sedentary behavior, and sleep quality experience.

Table 3. Performance of the truth interference methods using a ground truth data set of 9000 labeled tweets: 4000 physical activity, 2000 sedentary behavior, and 3000 sleep quality tweets. The top 4 rows of each PASS (physical activity, sedentary behavior, and sleep quality) category represent the results of the applied unsupervised truth inference models.

Tweets and method	Precision (%)		Recall (%)		F1 (%)		AUC_{PR}[a] (%)	
	Multiclass	Binary	Multiclass	Binary	Multiclass	Binary	Multiclass	Binary
Physical activity								
MV[b]	72	85	70	*85*[c]	71	84	56	85
DS[d]	74	85	68	*85*	70	84	54	85
GLAD[e]	73	84	70	84	71	83	57	84
RY[f]	74	85	68	*85*	70	84	54	84
LR[g]	74	85	*75*	*85*	*74*	*85*	*61*	87
KNN[h]	74	85	74	*85*	73	84	60	*88*
SVM[i]	72	*86*	73	*85*	73	*85*	*61*	*88*
RF[j]	73	85	74	84	73	*85*	60	87
XGBoost	72	81	72	81	71	81	58	83
DL_{meta}[k]	*79*	84	68	84	73	84	60	78
$DL_{text\_and\_meta}$	78	84	70	84	73	84	60	78
Sedentary behavior								
MV	71	82	68	82	68	82	54	80
DS	70	81	62	81	65	81	48	79
GLAD	71	79	68	79	68	79	54	77
RY	70	81	62	81	65	81	48	79
LR	72	*83*	*72*	*83*	70	*83*	*58*	*81*
KNN	71	82	71	82	67	82	56	80
SVM	73	*83*	*72*	*83*	70	*83*	*58*	*81*
RF	72	*83*	*72*	82	69	*83*	57	*81*
XGBoost	68	82	69	82	67	82	54	80
DL_{meta}	*78*	80	65	80	*71*	80	56	73
$DL_{text/meta}$	*78*	80	65	80	*71*	80	56	75
Sleep quality								
MV	78	*89*	74	*89*	75	*89*	61	87
DS	80	*89*	74	*89*	*77*	*89*	62	87
GLAD	79	85	75	85	76	85	62	82
RY	80	*89*	74	*89*	76	*89*	62	87
LR	76	88	*77*	87	*77*	88	64	88
KNN	76	*89*	*77*	*89*	*77*	*89*	63	*89*
SVM	76	88	*77*	88	*77*	88	64	88
RF	75	*89*	76	*89*	76	*89*	63	*89*
XGBoost	72	87	72	*89*	72	87	58	87
DL_{meta}	*82*	86	72	86	76	86	63	81
$DL_{text/meta}$	80	87	72	87	76	87	*65*	82

[a]AUC_{PR}: precision-recall area under the curve.

[b]MV: majority voting.

[c]Italicization indicates best performance for the metric and each PASS (physical activity, sedentary behavior, and sleep quality) category.

[d]DS: David and Skene.

[e]GLAD: generative model of labels, abilities, and difficulties.

[f]RY: Raykar algorithm.

[g]LR: logistic regression.

[h]KNN: K-nearest neighbors.

[i]SVM: support vector machine.

[j]RF: random forest.

[k]DL: deep learning.

Across all data sets, supervised models consistently performed better than unsupervised methods. This highlights the value of the context-sensitive information that was used as meta-information when training supervised models. However, on sleep quality, a data set with the same features and level of complexity as physical activity and sedentary behavior data sets, MV appears sufficient for the binary inference task, with supervised models providing little or no improvement.

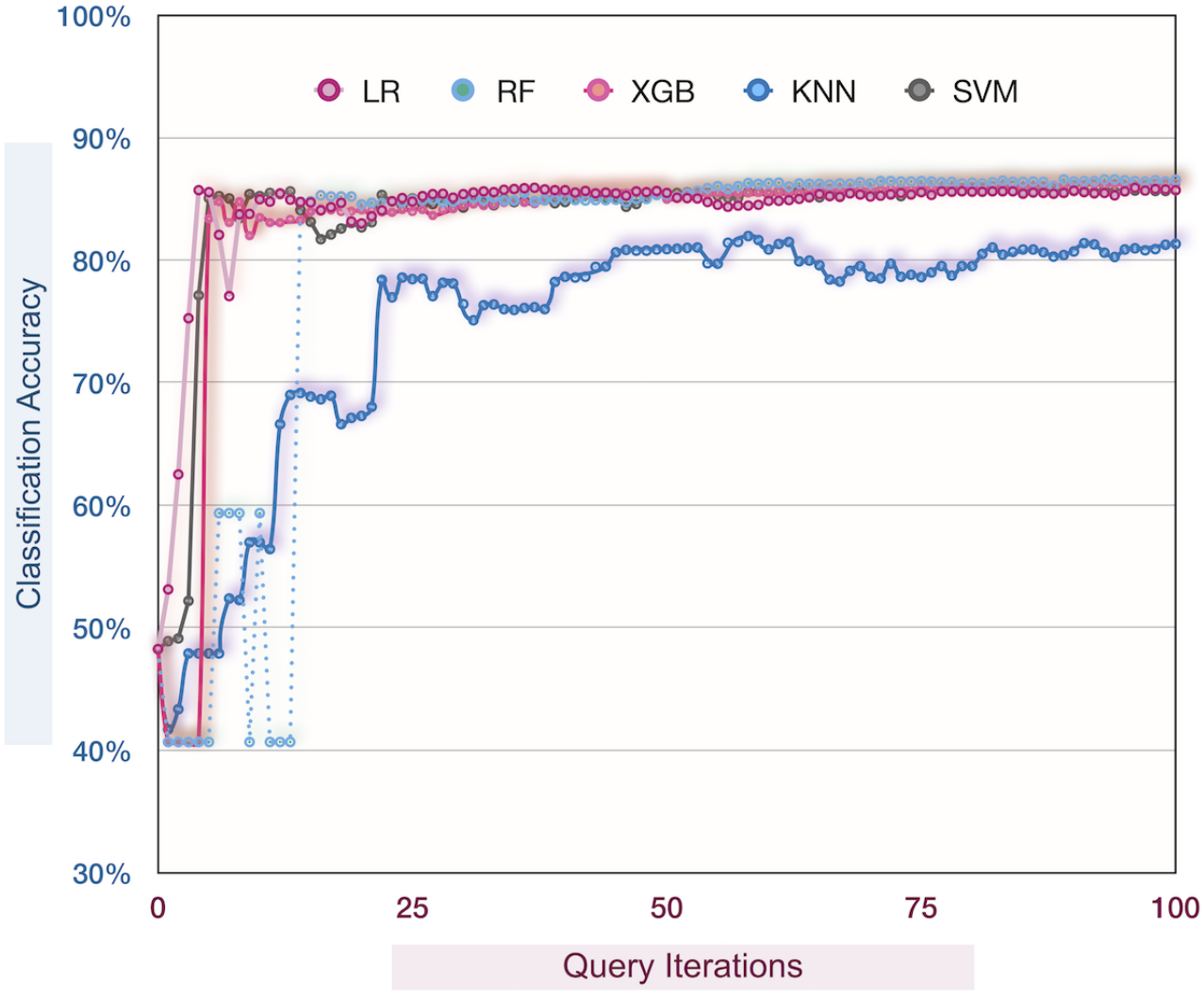
The hybrid CNN architecture did not provide any gain on either the unsupervised inference models or the supervised predictive models (ie, LR, KNN, SVM, RF, XGBoost, and DL_{meta}), and in some ways, underperformed them. It is possible that the LSTM stream could not capture the underlying dynamics of the features because of the inconsistencies between the poorly labeled tasks and the textual features.

Active Learning

To further explore the feasibility of correcting mislabeled samples, we used pool-based active learning [39] with uncertainty sampling. Pool-based active learning assumes that only a small set of data is labeled, and a large pool of data still needs to be labeled through an iterative learning process. All samples in the pool are queried based on an informativeness measure, which improves the learner's discrimination ability [40]. In this study, our learners were modeled to query the most ambivalent and uncertain samples. For example, for the binary label inference task, samples for which $p(\square = l \mid f) \approx 0.5$ are the most informative samples that may help detect mislabeled samples of the data set through different iterations. We used 5 different base learners with different architectures (ie, RF, LR, KNN, SVM, and XGBoost) with a batch size of 5 and queried the unlabeled pool through 100 iterations.

Our results show that, during the learning process, the accuracy of the classifiers generally increased, slightly degraded at some iterations, and stabilized around iteration 60 for KNN and iteration 20 for other classifiers (Figure 4). Although the active learners in this study could improve their predictive ability through a self-learning process, they failed to correct mislabeled samples and stabilized at performance scores lower than those of the offline learners discussed earlier (Table 3).

Figure 4. Incremental classification accuracy using pool-based active learning. KNN: K-nearest neighbors; LR: logistic regression; RF: random forest; SVM: support vector machine; XGB: XGBoost.



Discussion

Practical Recommendations

We start this section with some practical recommendations and guidelines on the use of AMT in specific and crowdsourcing in general for developing ML-based public health surveillance systems. Even under the assumption that more advanced artificial intelligence models, including pretrained models on general scope data sets and transfer-learning techniques, can cope with the poor quality of crow-generated labels, the guidelines provided in this study can still improve the implementation, design, and qualification of the crowd-labeling as well as the label inference processes. These guidelines are supported by the results described earlier and the findings and further analysis discussed in the rest of this section.

First, although the demographics of AMT workers are not available, we can still implement the crowdsourcing process in a way that accommodates a greater diversity of workers. A longitudinal labeling process, rather than one-time labeling, allows researchers to monitor the quality of the collected data over time, and mitigates the impact of spammers, irresponsible workers, and workers who are biased or mistake prone. Second, the overall quality of AMT workers can be context-sensitive and vary based on the type of labeling task. For example, the familiarity of the workers in the context of the tasks in the sleep quality data set, contrasting the broad context of physical activity and sedentary behavior concepts, resulted in higher data quality. Researchers should also be aware of the exclusion rate (eg, 5189/103,911, 4.99% in this study) and need to consider this when planning for their study's budget and design. Third, our study results show that consensus-based inference models that do not consider the task's features may not always be efficient for integrating crowdsourced labels and thus negatively impact the performance of ML models. Fourth, in addition to qualification requirements to filter crowdsourcing participants, sound and illustrative instruction is a less direct way to increase data quality. During the course of this project, we received nearly 70 emails from AMT workers, with most of them asked about scenarios that were mentioned in the instructions. This implies that the instruction changed their default understanding of the tasks, thereby improving the quality of the labels. Finally, when controlling the quality of workers using a qualification question, we recommend not informing the worker that this technique is being used, as they might guess the questions based on their simplicity.

Key Findings

Information Loss About Label Uncertainty

Despite all the alternative models developed in this study to improve the inference accuracy, there were still considerable discrepancies between workers and the truth labels. These disagreements may be attributable to the underlying uncertainty in the data. Although reducing uncertainty by collecting more labels from more workers might simplify the process of label inference, it limits the learning ability of ML models in modeling the inherent uncertainty of data and prevents them from recovering from the mistakes made early during the inference process [41].

Robustness of Inference Models

We observed from our inference results that, regardless of the type of the classification task, none of the 11 methods outperformed other methods across all data sets (Table 3). This indicates that inference methods are sensitive to data set characteristics. For example, the performance of all of the methods on the sleep quality data set is better than that of physical activity and sedentary behavior data sets, indicating the low robustness of these models against the task context.

The Importance of Task Features

Compared with supervised models that require a large volume of labeled data to integrate crowd-generated labels, using unsupervised inference models is simple and straightforward. However, this simplicity is gained through the cost of throwing away the contextual characteristics of tasks, which may sacrifice quality in context-sensitive scenarios. For example, the time that a tweet is posted during a day can contribute to the decision about its relevance to physical activity or sleep quality contexts. The importance of these characteristics was far more pronounced in the multiclass inference tasks than in the binary tasks (Table 3), suggesting the need for more complicated models when inferring the truth label of tasks with a high level of uncertainty.

The Effectiveness of Qualification Requirements

In this study, we used two levels of quality control: (1) through the task assignment process by accepting only workers with a master qualification and (2) through the design and implementation of the tasks by adding a qualification question to our HITs and iteratively observing workers' performance based on their answer to this question. Our results show that even though defining these requirements improved the quality of crowd-generated labels to a great extent, 12.45% (498/4000), 13.3% (266/2000), and 7.7% (231/3000) of physical activity, sedentary behavior, and sleep quality tweets, respectively, were still mislabeled by all three workers, regardless of their context or complexity level, indicating the need for further quality assessment of crowdsourced data. These mislabeled samples were not misclassified due to sample uncertainty or difficulty, and our further analysis shows that they were not informative enough (ie, prediction scores) to improve the performance of predictive models through the iterative process of active learning (Figure S4 in Multimedia Appendix 1). Considering the sparsity of the (workers and tasks) matrix in large-scale crowdsourcing tasks, distinguishing irresponsible workers and removing their impact is a challenging task that should be carefully considered when training ML models based on crowd-labeled data. A sample list of low-quality labels for all the PASS categories is provided in Figure S5 in Multimedia Appendix 1.

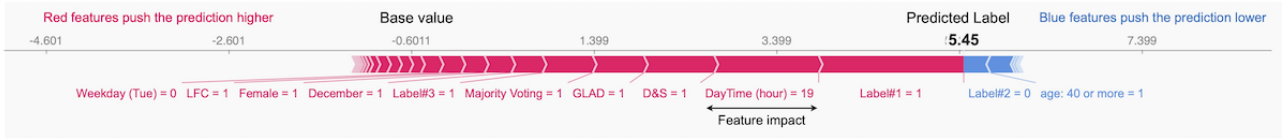
The Impact of Crowd-Generated Labels on the Performance of Predictive Models

To further investigate the reliability of using crowdsourcing for developing ML models, we used bidirectional encoder representations from transformers [42] (ie, bert-base-uncased); a transformer-based model with 12-layer, 768 hidden units, 12 heads; and 110M parameters as a contextual input to our DL model, to classify 4000 physical activity tweets, using our binary truth labels and crowd-generated labels. We used the labels inferred by SVM for the crowd-generated labels, as it outperformed other models on the physical activity data set (Table 3). Interestingly, the model that was trained on our ground truth data set outperformed the crowd-labeled data set on all performance metrics by at least 8% (eg, crowd-labeled: AUC_{PR} of 72%; expert-labeled: AUC_{PR} of 82%). This indicates the importance of the quality of crowd-generated labels in developing ML models designed for decision-making purposes, such as public health surveillance decisions.

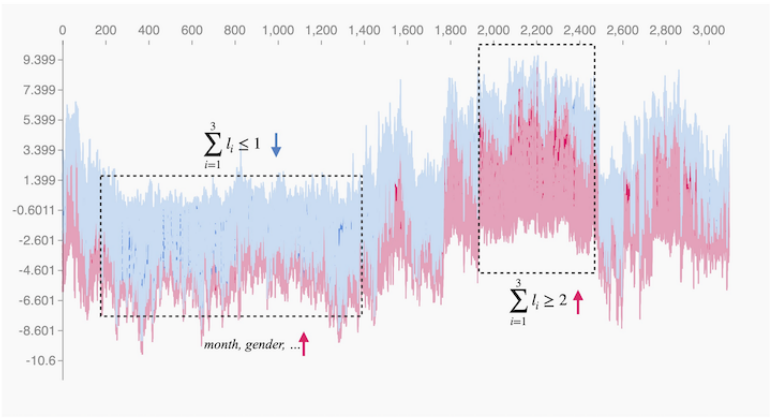
Label Prediction Explanation

To interpret the results of our predictive models in terms of the individual contribution of each feature to the prediction results, we used SHAP [23,43]. SHAP calculates the local, instead of global, feature importance for each sample of the data set, which mitigates the risks associated with inconsistency problems in other feature importance techniques. Figure 5A illustrates the interpretation of the prediction using XGBoost on a randomly selected sample of the physical activity data set using SHAP. The red arrows show the features that contribute to the increase, and the blue arrows represent features that contribute to the decrease in the prediction. The width of each arrow indicates the height of its impact. From this example, we can see that l_1=1 and daytime=7pm have the most positive impact on the predicted label, whereas l_2=0 and age ≥40 has the most negative impact.

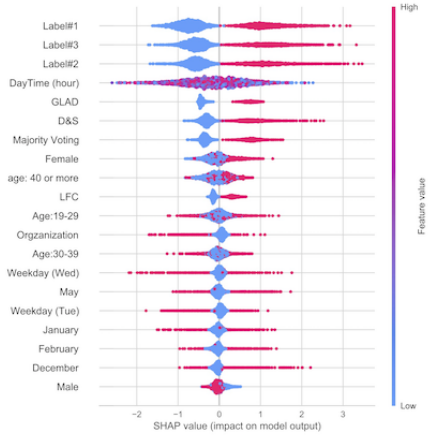
Figure 5. The estimated impact of each piece of meta-information on XGBoost when predicting the truth label. Age is in years. D&S: David and Skene; GLAD: generative model of labels, abilities, and difficulties; LFC: Learning from Crowds (Raykar algorithm); SHAP: Shapley additive explanations.



(a) Force plot for one sample of the dataset



(b) Stacked SHAP diagram



(c) Summary plot

We further used Shapely values to cluster our data set based on the explanation similarity of samples, using hierarchical agglomerative clustering (Figure 5B). From this figure, we can see that the crowdsourced labels are the most influential features in grouping the samples in our data set. The highlighted areas in this diagram show the samples that have similar force plots, implying the dominant and similar contribution of these features across the physical activity data set.

Using the additive nature of Shapley values, we integrated all the local feature values for each data point and calculated the global contribution (I) of each feature. Considering $\varphi_{j,i}$ as the Shapley value of feature j for sample i, we can calculate the global importance of this feature as $I_j = \sum_{i=1}^{N} |\varphi_{j,i}|$. Figure 5C shows the combination of feature importance (y-axis) and feature effects (colored points) for the most influential features, ordered based on their importance. This plot shows that crowdsourced labels (l_1, l_2, and l_3), followed by *daytime*, the results of the *inference models*, and *gender* have the greatest impact on the decision-making of XGBoost. From these results, which are extendable to the other predictive models developed in this study, it can be inferred that regardless of the complexity and the architecture of the predictive models, the crowd-generated labels are the factors that most influence predictive models' prediction. Although meta-information such as *daytime* and *gender* are among the most contributing features (Figure 5C), they still cannot compete with the crowd-generated labels in most of the samples. This can explain the vulnerability of our ML and DL models to the noisy labels of the data set.

To triangulate the dominant impact of the crowdsourced labels, we excluded all the samples for which $\sum_{i=1}^{3} l_i \leq 1$ or from our data set for both supervised and unsupervised techniques and achieved an F_1 score of approximately 99%. This implies that inferring the truth label of crowdsourced data highly depends on the quality of the collected data from the crowd, and even advanced and complex predictive models might not be able to compensate for the poor quality of these data.

Limitations

This study has several limitations. First, the compensation paid to the workers could impact the quality of the collected labels, and consequently, the evaluation results of this study. Workers may show a higher quality in exchange for higher payments. To investigate this, during the course of the project, we increased HITs' reward from US $0.03 to US $0.05 and did not notice any significant changes in quality. However, this is still debatable and requires further investigation.

Second, to develop the supervised models, we assumed that all the tasks share the same level of complexity, whereas in reality, some examples are more difficult than others. For example, labeling "I can't sleep" to a self-reported sleep problem is more straightforward than labeling "I'm kind of envious of anyone who is able to fall asleep before 2am." We attempted to address this by incorporating inherent task difficulties in the prediction models by developing a hybrid CNN model. However, crowd-generated labels dominated other features of our data set, which had the greatest impact on their inference decisions. Building crowdsourcing models sensitive to the complexity of tasks to allocate more resources (workers) to more difficult tasks is a worthwhile direction for future research.

Third, the way we designed and presented the HITs on AMT could impact the performance of workers in various ways. Considering the central role of people in maximizing the benefits of crowdsourcing services, human factors should be considered when designing crowdsourcing tasks [41]. To address this, we added succinct, precise, and demonstrative instructions to each task and explained each label with an illustrative example (eg,

Figure S6 in Multimedia Appendix 1). In addition, through different iterations of data collection, we tweaked the design, presentation, and instructions to ensure that we met the basic usability requirements of task design and presentation.

Fourth, we defined workers' qualifications based only on their historical performance in completing HITs across AMT (ie, master qualification). Although this provided some degree of quality control on the collected labels, alternative qualification requirements such as workers' education, work background, and language could have also impacted our study results. To further study the role of qualification filtering, we pilot-tested the labeling process without any qualification requirements for 4500 physical activity tasks. These tasks were completed in <12 hours with a consistency score (*LC*) of <0.5, implying the importance of workers' quality in developing crowd-labeled intelligent systems.

Fifth, various physical activities, based on their energy requirements in metabolic equivalents (METs), can be categorized into different movement behaviors, such as light (1.6-2.9 METs), moderate (3-5.9 METs), and vigorous (≥6 METs) [44]. However, as the details provided by social media data may not be enough to calculate the MET values, in this study, we only used general terms related to physical activity (eg, physical fitness, exercise, household, sports, or occupational activities) to filter and form the physical activity subset. To ensure that the lists of contextual terms for filtering all the PASS categories are comprehensive enough, in addition to domain-specific ontologies and WordNet [45], we used NLP techniques (eg, topic modeling, language modeling, and lexical analysis) to detect latent word patterns that can be used to identify PASS-related contexts in unstructured text. However, with no impact on the methodology and results of this study, both data collection and population biases (inherent in social media data) should be considered when discussing the data set used for this study.

Despite these limitations, our study is one of the first to rigorously investigate the challenges of using crowdsourcing to develop ML-based public health surveillance systems. Our findings support the argument that crowdsourcing, despite its low cost and short turnaround time, yields noisier data than in-house labeling. On the flip side, crowdsourcing can reduce annotation bias by involving a more diverse set of annotators [41]. This diversity, supported by the diversity of AMT workers [46], is highly beneficial to subjective labeling tasks, such as detecting a sedentary behavior based on a short text, which highly depends on the worker's understanding of sedentary lifestyles.

The results of this study may inspire future research to investigate and evaluate the application of crowdsourcing for the development of ML-based digital public health surveillance systems deployed and used in national surveillance decision-making. As the potential for success of ML-based digital public health surveillance relies on robust and reliable data sets, a sensitivity analysis of health-related incidents detected by ML-based surveillance models trained on crowd-generated labels versus relevant national datasets is required to ascertain this potential. Moreover, to assess whether our conclusions are sensitive to the background and expertise of participants, further investigation is required using a cohort of experts who are familiar with the public health context under study. Likewise, to untangle the effect of task context and the quality of the crow-generated labels, replicating the approach adopted in this study using other domains, including other public health domains, remains a future work. Finally, as there is a chance that the quality of the crowd-generated labels is subject to the compensation amount, confounded by the socioeconomic characteristics of the participant cohort, future investigations are required to calibrate the results of this study considering these factors.

Acknowledgments

This work was supported by a postdoctoral scholarship from the Libin Cardiovascular Institute and the Cumming School of Medicine, University of Calgary. This work was also supported by a Discovery Grant from the Natural Sciences and Engineering Research Council of Canada (RGPIN-2014-04743). The Public Health Agency of Canada funded the Amazon Mechanical Turk costs. The funders of the study had no role in the study design, data collection and analysis, interpretation of results, and preparation of the manuscript.

Authors' Contributions

ZSHA was responsible for data collection and curation, model development, data analysis, and visualization, and wrote the paper. GPB and WT reviewed the paper and provided comments. JL conceived and designed the study and revised the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Additional figures that describe the Amazon Mechanical Turk labeling task, predictive model performance, and incorrectly labeled tweets in more detail.
[PDF File (Adobe PDF File), 1434 KB - jmir_v24i1e28749_app1.pdf]

References

1. Mavragani A. Infodemiology and infoveillance: scoping review. J Med Internet Res 2020 Apr 28;22(4):e16206 [FREE Full text] [doi: 10.2196/16206] [Medline: 32310818]
2. Aiello AE, Renson A, Zivich PN. Social media- and internet-based disease surveillance for public health. Annu Rev Public Health 2020 Apr 02;41(1):101-118 [FREE Full text] [doi: 10.1146/annurev-publhealth-040119-094402] [Medline: 31905322]
3. Sinnenberg L, Buttenheim AM, Padrez K, Mancheno C, Ungar L, Merchant RM. Twitter as a Tool for Health Research: A Systematic Review. Am J Public Health 2017 Jan;107(1):e1-e8. [doi: 10.2105/AJPH.2016.303512] [Medline: 27854532]
4. Bernardo TM, Rajic A, Young I, Robiadek K, Pham MT, Funk JA. Scoping review on search queries and social media for disease surveillance: a chronology of innovation. J Med Internet Res 2013 Jul 18;15(7):e147 [FREE Full text] [doi: 10.2196/jmir.2740] [Medline: 23896182]
5. Hossain L, Kam D, Kong F, Wigand RT, Bossomaier T. Social media in Ebola outbreak. Epidemiol Infect 2016 Jul;144(10):2136-2143. [doi: 10.1017/S095026881600039X] [Medline: 26939535]
6. Lardon J, Abdellaoui R, Bellet F, Asfari H, Souvignet J, Texier N, et al. Adverse drug reaction identification and extraction in social media: a scoping review. J Med Internet Res 2015 Jul 10;17(7):e171 [FREE Full text] [doi: 10.2196/jmir.4304] [Medline: 26163365]
7. Hu H, Phan N, Chun SA, Geller J, Vo H, Ye X, et al. An insight analysis and detection of drug-abuse risk behavior on Twitter with self-taught deep learning. Comput Soc Netw 2019 Nov 06;6(1):10. [doi: 10.1186/s40649-019-0071-4]
8. Cavallo DN, Tate DF, Ries AV, Brown JD, DeVellis RF, Ammerman AS. A social media-based physical activity intervention: a randomized controlled trial. Am J Prev Med 2012 Nov;43(5):527-532 [FREE Full text] [doi: 10.1016/j.amepre.2012.07.019] [Medline: 23079176]
9. Dunn AG, Mandl KD, Coiera E. Social media interventions for precision public health: promises and risks. NPJ Digit Med 2018 Sep 19;1(1):1-4 [FREE Full text] [doi: 10.1038/s41746-018-0054-0] [Medline: 30854472]
10. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. Health Inf Sci Syst 2014 Feb 7;2(1):3 [FREE Full text] [doi: 10.1186/2047-2501-2-3] [Medline: 25825667]
11. Shakeri Hossein Abad Z, Kline A, Sultana M, Noaeen M, Nurmambetova E, Lucini F, et al. Digital public health surveillance: a systematic scoping review. NPJ Digit Med 2021 Mar 03;4(1):41 [FREE Full text] [doi: 10.1038/s41746-021-00407-6] [Medline: 33658681]
12. Paolacci G, Chandler J, Ipeirotis P. Running experiments on Amazon Mechanical Turk. Judgm Dec Mak 2010;5(5):1-9 [FREE Full text]
13. Peer E, Brandimarte L, Samat S, Acquisti A. Beyond the Turk: alternative platforms for crowdsourcing behavioral research. J Experiment Soc Psychol 2017 May;70:153-163. [doi: 10.1016/j.jesp.2017.01.006]
14. Brabham DC, Ribisl KM, Kirchner TR, Bernhardt JM. Crowdsourcing applications for public health. Am J Prev Med 2014 Feb;46(2):179-187. [doi: 10.1016/j.amepre.2013.10.016] [Medline: 24439353]
15. Kim SJ, Marsch LA, Hancock JT, Das AK. Scaling up research on drug abuse and addiction through social media big data. J Med Internet Res 2017 Oct 31;19(10):e353 [FREE Full text] [doi: 10.2196/jmir.6426] [Medline: 29089287]
16. Lu W, Guttentag A, Elbel B, Kiszko K, Abrams C, Kirchner TR. Crowdsourcing for food purchase receipt annotation via Amazon Mechanical Turk: a feasibility study. J Med Internet Res 2019 Apr 05;21(4):e12047 [FREE Full text] [doi: 10.2196/12047] [Medline: 30950801]
17. Ayers JW, Leas EC, Allem J, Benton A, Dredze M, Althouse BM, et al. Why do people use electronic nicotine delivery systems (electronic cigarettes)? A content analysis of Twitter, 2012-2015. PLoS One 2017 Mar 1;12(3):e0170702 [FREE Full text] [doi: 10.1371/journal.pone.0170702] [Medline: 28248987]
18. Yin Z, Fabbri D, Rosenbloom ST, Malin B. A scalable framework to detect personal health mentions on Twitter. J Med Internet Res 2015 Jun 05;17(6):e138 [FREE Full text] [doi: 10.2196/jmir.4305] [Medline: 26048075]
19. McIver DJ, Hawkins JB, Chunara R, Chatterjee AK, Bhandari A, Fitzgerald TP, et al. Characterizing sleep issues using Twitter. J Med Internet Res 2015 Jun 08;17(6):e140 [FREE Full text] [doi: 10.2196/jmir.4476] [Medline: 26054530]
20. Reece AG, Reagan AJ, Lix KL, Dodds PS, Danforth CM, Langer EJ. Forecasting the onset and course of mental illness with Twitter data. Sci Rep 2017 Oct 11;7(1):13006 [FREE Full text] [doi: 10.1038/s41598-017-12961-9] [Medline: 29021528]
21. Adrover C, Bodnar T, Huang Z, Telenti A, Salathé M. Identifying adverse effects of HIV drug treatment and associated sentiments using Twitter. JMIR Public Health Surveill 2015 Jul 27;1(2):e7 [FREE Full text] [doi: 10.2196/publichealth.4488] [Medline: 27227141]
22. Peer E, Vosgerau J, Acquisti A. Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. Behav Res 2013 Dec 20;46(4):1023-1031. [doi: 10.3758/s13428-013-0434-y]
23. Lundberg S, Lee S. A unified approach to interpreting model predictions - advances in neural information processing systems. In: Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS 2017). 2017 Presented at: 31st Conference on Neural Information Processing Systems (NIPS 2017); 2017; Long Beach, CA, USA p. 4765-4774 URL: https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

24. Wang Z, Hale S, Adelani D, Grabowicz P, Hartman T, Flöck F, et al. Demographic inference and representative population estimates from multilingual social media data. In: Proceedings of the World Wide Web Conference. 2019 Presented at: WWW '19: The Web Conference; May 13 - 17, 2019; San Francisco CA USA p. 2056-2067. [doi: 10.1145/3308558.3313684]
25. Shakeri Hossein Abad Z, Butler GP, Thompson W, Lee J. Physical activity, sedentary behaviour, and sleep on Twitter: a multicountry and fully labelled dataset for public health surveillance research. JMIR Preprints. Preprint posted online July 23, 2021 [FREE Full text] [doi: 10.2196/32355]
26. Zheng Y, Li G, Li Y, Shan C, Cheng R. Truth inference in crowdsourcing: is the problem solved? Proc VLDB Endow 2017 Jan;10(5):541-552. [doi: 10.14778/3055540.3055547]
27. Sheshadri A, Lease M. Square: a benchmark for research on computing crowd consensus. In: Proceedings of the First AAAI Conference on Human Computation and Crowdsourcing. 2013 Presented at: First AAAI Conference on Human Computation and Crowdsourcing; November 7-9, 2013; Palm Springs, California, USA p. 156-164 URL: https://ojs.aaai.org/index.php/HCOMP/article/view/13088
28. Dawid AP, Skene AM. Maximum likelihood estimation of observer error-rates using the EM algorithm. Appl Stat 1979;28(1):20. [doi: 10.2307/2346806]
29. Ipeirotis P, Provost F, Wang J. Quality management on Amazon Mechanical Turk. In: Proceedings of the ACM SIGKDD Workshop on Human Computation. 2010 Presented at: KDD '10: The 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; July 25, 2010; Washington DC p. 64-67. [doi: 10.1145/1837885.1837906]
30. Whitehill J, Wu T, Bergsma J, Movellan J, Ruvolo P. Whose vote should count more: optimal integration of labels from labelers of unknown expertise. In: Proceedings of the 23rd Annual Conference on Neural Information Processing Systems. 2009 Presented at: 23rd Annual Conference on Neural Information Processing Systems; December 7-10, 2009; Vancouver, British Columbia Canada p. 2035-2043 URL: https://papers.nips.cc/paper/2009/hash/f899139df5e1059396431415e770c6dd-Abstract.html
31. Raykar V, Yu S, Zhao L, Valadez G, Florin C, Bogoni L, et al. Learning from crowds. J Mach Learn Res 2010;11(43):1297-1322 [FREE Full text]
32. Pennington J, Socher R, Manning C. Glove: Global vectors for word representation. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2014 Presented at: Conference on Empirical Methods in Natural Language Processing (EMNLP); October 2014; Doha, Qatar p. 1532-1543. [doi: 10.3115/v1/d14-1162]
33. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
34. Snoek J, Larochelle H, Adams R. Practical bayesian optimization of machine learning algorithms. In: Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS 2012). 2012 Presented at: Advances in Neural Information Processing Systems 25 (NIPS 2012); December 3-6, 2012; Lake Tahoe, Nevada, USA p. 2951-2959 URL: https://proceedings.neurips.cc/paper/2012/hash/05311655a15b75fab86956663e1819cd-Abstract.html
35. Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation. 2016 Presented at: 12th USENIX Conference on Operating Systems Design and Implementation; November 2 - 4, 2016; Savannah GA USA p. 265-283. [doi: 10.5555/3026877.3026899]
36. Chollet F. Keras: the python deep learning library. Astrophysics Source Code Library. 2018. URL: https://ui.adsabs.harvard.edu/abs/2018ascl.soft06022C/abstract [accessed 2021-12-11]
37. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. J Mach Learn Res 2011;12(10):2825-2830 [FREE Full text]
38. CrowdSourcing-for-Digital-Public-Health-Surveillance. GitHub. URL: https://github.com/data-intelligence-for-health-lab/CrowdSourcing-for-Digital-Public-Health-Surveillance [accessed 2021-12-16]
39. Lewis D, Gale W. A Sequential Algorithm for Training Text Classifiers. London, UK: Springer; 1994:3-12.
40. Laws F, Scheible C, Schütze H. Active learning with Amazon Mechanical Turk. In: Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. 2011 Presented at: Conference on Empirical Methods in Natural Language Processing; July 2011; Edinburgh, Scotland, UK p. 1546-1556 URL: https://aclanthology.org/D11-1143/
41. Lease M. On quality control and machine learning in crowdsourcing. Hum Comput 2011;11(11):1-6 [FREE Full text]
42. Devlin J, Chang M, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding. arXiv. 2018. URL: https://arxiv.org/abs/1810.04805 [accessed 2021-12-11]
43. Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable ai for trees. Nat Mach Intell 2020 Jan 17;2(1):56-67 [FREE Full text] [doi: 10.1038/s42256-019-0138-9] [Medline: 32607472]
44. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Rep 1985;100(2):126-131 [FREE Full text] [Medline: 3920711]
45. Miller GA. WordNet: a lexical database for English. Commun ACM 1995 Nov;38(11):39-41. [doi: 10.1145/219717.219748]
46. Difallah D, Filatova E, Ipeirotis P. Demographics and dynamics of mechanical Turk workers. In: Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining. 2018 Presented at: WSDM 2018: The Eleventh ACM

International Conference on Web Search and Data Mining; February 5 - 9, 2018; Marina Del Rey CA USA p. 135-143. [doi: 10.1145/3159652.3159661]

Abbreviations

AMT: Amazon Mechanical Turk
AUCPR: precision-recall area under the curve
CNN: convolutional neural network
DL: deep learning
DS: David and Skene
EM: expectation–maximization
GLAD: generative model of labels, abilities, and difficulties
HIT: Human Intelligence Task
KNN: K-nearest neighbors
LR: logistic regression
LSTM: long short-term memory
MET: metabolic equivalent
ML: machine learning
MV: majority voting
NLP: natural language processing
PASS: physical activity, sedentary behavior, and sleep quality
ReLU: Rectified Linear Unit
RY: Raykar algorithm
SHAP: Shapley Additive Explanations
SMOTE: Synthetic Minority Oversampling Technique-Nominal Continuous
SVM: support vector machine

Edited by R Kukafka; submitted 13.03.21; peer-reviewed by R Xu, A Das; comments to author 12.06.21; revised version received 05.07.21; accepted 15.11.21; published 18.01.22.

Please cite as:
Shakeri Hossein Abad Z, Butler GP, Thompson W, Lee J
Crowdsourcing for Machine Learning in Public Health Surveillance: Lessons Learned From Amazon Mechanical Turk
J Med Internet Res 2022;24(1):e28749
URL: https://www.jmir.org/2022/1/e28749
doi:10.2196/28749
PMID:35040794

Original Paper

A New Method to Extract Health-Related Quality of Life Data From Social Media Testimonies: Algorithm Development and Validation

Simon Renner[1*], PharmD; Tom Marty[1*], PharmD; Mickaïl Khadhar[1*], MSc; Pierre Foulquié[1], MSc; Paméla Voillot[1], MSc; Adel Mebarki[1], MSc; Ilaria Montagni[2], DPhil; Nathalie Texier[1], PharmD; Stéphane Schück[1], MD

[1]Kap Code, Paris, France

[2]Bordeaux Population Health Research Center, UMR 1219, Bordeaux University, Inserm, Bordeaux, France

[*]these authors contributed equally

Corresponding Author:
Simon Renner, PharmD
Kap Code
4 Rue de Cléry
Paris, 75002
France
Phone: 33 9 72 60 57 63
Email: simon.renner@kapcode.fr

Abstract

Background: Monitoring social media has been shown to be a useful means to capture patients' opinions and feelings about medical issues, ranging from diseases to treatments. Health-related quality of life (HRQoL) is a useful indicator of overall patients' health, which can be captured online.

Objective: This study aimed to describe a social media listening algorithm able to detect the impact of diseases or treatments on specific dimensions of HRQoL based on posts written by patients in social media and forums.

Methods: Using a web crawler, 19 forums in France were harvested, and messages related to patients' experience with disease or treatment were specifically collected. The SF-36 (Short Form Health Survey) and EQ-5D (Euro Quality of Life 5 Dimensions) HRQoL surveys were mixed and adapted for a tailored social media listening system. This was carried out to better capture the variety of expression on social media, resulting in 5 dimensions of the HRQoL, which are physical, psychological, activity-based, social, and financial. Models were trained using cross-validation and hyperparameter optimization. Oversampling was used to increase the infrequent dimension: after annotation, SMOTE (synthetic minority oversampling technique) was used to balance the proportions of the dimensions among messages.

Results: The training set was composed of 1399 messages, randomly taken from a batch of 20,000 health-related messages coming from forums. The algorithm was able to detect a general impact on HRQoL (sensitivity of 0.83 and specificity of 0.74), a physical impact (0.67 and 0.76), a psychic impact (0.82 and 0.60), an activity-related impact (0.73 and 0.78), a relational impact (0.73 and 0.70), and a financial impact (0.79 and 0.74).

Conclusions: The development of an innovative method to extract health data from social media as real time assessment of patients' HRQoL is useful to a patient-centered medical care. As a source of real-world data, social media provide a complementary point of view to understand patients' concerns and unmet needs, as well as shedding light on how diseases and treatments can be a burden in their daily lives.

(J Med Internet Res 2022;24(1):e31528) doi:10.2196/31528

KEYWORDS

health-related quality of life; social media use; measures; real world; natural language processing; social media; NLP; infoveillance; quality of life; digital health; social listening

Introduction

Most people use the internet regularly to research and discuss health-related topics. Patients give and receive advice on their diseases and treatments in online forums and social media platforms [1]. These messages are massive, continuously generated, and easy to access [2]. This type of information is direct, genuine, and authentic, offering access to new real-world

data, which can facilitate the understanding of patients' perspectives. As the internet offers anonymity, patients talk about their fears and concerns and share details about their diseases and treatments, which can inform health public authorities, pharmaceutical companies, and other health professionals and institutions [3]. Thus, social media are a large and diverse source of information nurtured by continuous exchanges and interactions, ranging from commenting on posts to sharing of opinions.

The World Health Organization defines quality of life (QoL) as individuals' perception of their place in life in the context of the culture and the value system in which they live, as well as in relation to their objectives, expectations, standards, and concerns. This is a broad conceptual field, encompassing, in a complex way, a person's physical health, psychological state, level of independence, social relationships, personal beliefs, and relationship with the specificities of the surrounding environment [4]. When the study of QoL is restricted to health-related effects, one can refer to them as health-related quality of life (HRQoL) [5]. Therefore, HRQoL is a multidimensional concept focusing on the impact health and diseases have on QoL [6,7]. This concept is mainly used in epidemiology and cost-effectiveness analysis [8].

Several instruments have been developed to quantitatively measure individuals' HRQoL [9]. Among them, the EQ-5D (Euro Quality of Life 5 Dimensions) and SF-36 (Short Form Health Survey) have been used in medical practice for more than 20 years [10,11]. They are designed to be self-completed by patients. Nonetheless, these surveys are not adapted to the amount of qualitative information on QoL contained within the free speech and various testimonies of patients' populations on social media.

It has been suggested that the measurement of HRQoL can benefit from machine-driven, quantitative analysis of patient-generated data, which expands hypothesis testing based on patient input regarding disease experience, lifestyle preferences, functioning, and more [12]. Opinions and advice shared on social media can provide insights on HRQoL directly from patients in real-life conditions [13].

Social media listening is the collection and interpretation of all patients' social media conversations, which can help discover what really impacts patients' lives [14]. Social media listening aggregates large amounts of unstructured patient-centered data points to identify behavioral patterns and obtain medical insights without infringing privacy policy and personal rights. Social media listening uses text mining and the natural language processing (NLP) approach as an algorithmic toolbox for identifying and managing texts of interest [15].

Against this background, the objective of this study was to develop an algorithm that is able to detect and measure the mentions of impact of diseases and treatments on 5 HRQoL dimensions in patient's testimonies through the scope of social media listening.

Methods

This study was conducted through several main steps: QoL definition, literature review, data extraction and manual treatment, annotation, preprocessing and feature engineering, modeling, and statistical analysis.

Health-Related Quality of Life

The European Knowledge Society on Quality and HRQoL has compared the many definitions of HRQoL and discussed the existing confusions between health, QoL, HRQoL, and well-being [8]. The EQ-5D tool is recommended by the French public institute that regulates recommendations toward health products, their uses, and efficacy measurement (Haute Autorité de Santé [16,17]). The SF-36 is another validated generic medical survey investigating HRQoL, broadly used by practitioners for years. Three dimensions are always at the heart of the definitions or surveys: physical, psychological, and social. However, exploring HRQoL (especially on social media, with the spontaneous discussions of patients) can shed light on other views and aspects of an individual, including economic, spiritual, or even political matters. Therefore, in addition to the 3 constant dimensions (physical, psychological, and social dimension), 2 more dimensions were added to the methodology for their important role in one's life, which can especially be impacted in the case of diseases. The dimension of generic activity is unavoidable in one's life and can be limited in some health states, from taking a shower to professional activities; therefore, the aim of analyzing a 4th dimension is to detect mentions of impact on patients' activity and autonomy, which are complementary to the physical dimension that focuses on body impairments. The 5th dimension is the financial one; according to the definitions developed by the European Knowledge Society on Quality and HRQoL, economic and personal finances are important contextual factors to patients [8]. Some can encounter bad or no insurances toward treatment costs or must pay for parallel cares or products that are not covered by their insurance. Patients can express specific health expenses or the necessity to have a specific budget because of their disease; therefore, the financial dimension covers this relation between health state and the impact on one's finances as expressed by patients in their messages.

A previous work by Cotté et al [13] showed that posts from social media could be used to assess the impact of a disease or a treatment on HRQoL. This study, focused on the narratives of cancer patients treated with immunotherapies, highlighted that posts from patients could provide additional information on HRQoL to conventional QoL measurement instruments (ie, QLQ-C30 [Quality of Life Questionnaire] and FACT-G [Functional Assessment of Cancer Therapy—General]).

Literature Review

We searched on PubMed and Google Scholar for articles responding to the following keywords: (natural language processing[MeSH Terms] OR processing natural language[MeSH Terms]) AND (quality of life[MeSH Terms] OR health related quality of life[MeSH Terms] OR healthrelated quality of life[MeSH Terms] OR HRQOL[MeSH Terms] OR cost of illness[MeSH Terms] OR disease burden[MeSH Terms]

OR sickness impact profile[MeSH Terms]). The selected results were based on NLP, social media, patients' messages, QoL, diseases, and side effects. About 40 articles were found and used with the aim of establishing the best method and modeling to adopt (Multimedia Appendix 1). A focus was made on articles that developed machine learning techniques over neural network because of their lower cost in resources and correspondence with our database. The takeaway from literature review is that some machine learning methods, tools, or approaches were highlighted for their good performance in the literature review, such as Naive Bayes, Max Entropy, Decision Tree (10 folds), and MaxVote. AdaBoost has been used for its performance boost in the learning phases. The overall performances showed that the combination of a binary classifier was better than the use of only 1 predictive model. Concerning the supervised learning for text classification, a stacked generalization method, such as SVM-L (support vector machine light), SVM-R (support vector machine regression), GBDT (gradient boosting decision tree), Unigram, bigrams, POS (part of speech), and TF-IDF (term frequency-inverse document frequency), has proven interesting for obtaining state-of-the-art results [18].

Data Sources and Manual Treatment

The sources of data were 19 online general or health-related community forums in France, which are as follows: Atoute [19], Doctissimo [20], AuFeminin [21], Journal des femmes [22], Psychoactif [23], Forum.hardware [24], Lesimpatientes [25], Laxophobie [26], Magic maman [27], thyroide [28], forum ado/public.fr [29], Onmeda [30], Psychologies [31], MeaMedica [32], Futura-sciences [33], Allodocteurs [34], Vulgaris Medical [35], Lymphome espoir [36], and Maman pour la vie [37]. Facebook and Twitter were not included because tweets are limited to 240 characters, which limited the probability of disease history development and impact testimonies. Facebook was also discarded for data privacy questions and difficulties of access. Messages were extracted using a web crawler technology [38,39]. Health-related messages were selected based on a named entity recognition (NER) module. NER is a process where a sentence is parsed through to find entities (names, organizations, locations, and quantities). The NER module was used here to identify drug or disease mentions using an approximate matching algorithm. These messages were then preprocessed and stored. The metadata extracted along with the text were the date and hour of post.

Raw data sets were composed of randomly selected health-related messages according to the presence of treatment or disease in it. Preprocessing of the extracted data included a code attribution to every message as identifier, the detection of sentences, normalization, and deduplication; since the extracted data were unstructured, this was a necessary first step to process patients' posts.

Annotation

The corpus (n=1399 posts), with 1000 (71%) posts with disease mentions and 399 (29%) posts with treatment mentions, was first manually annotated. Manual annotation was performed by 2 individuals: a health-specialized data scientist and a health care professional specialized in social media listening, both sensitized and trained about the medical field of QoL, following guidelines in accordance with the methodology of HRQoL. The 2 annotators' profiles worked in synergy in the approach of data annotation with a medical finality. Medical insight toward patients' testimonies was brought by one of the annotators, with an expert eye toward the variables to be included in the future models by the other. The 1399 health-related messages extracted from forums were split into 2 sets for labelling; respectively, 900 and 499 messages for the 2 annotators. The aim of this step was to classify the messages according to 5 specific dimensions corresponding to 5 different types of impact: physical, psychic, activity-related, relational, and financial. The labels data were either "not impacted" or "impacted." If "impacted," the concerned dimensions were characterized through annotation, and the patients' expressions of the said impact were extracted. This collection allowed the identification of specific features for each dimension being impacted, capturing the patients' vocabulary when mentioning the impact. To evaluate the annotation homogeneity, a subset of 100 messages coming from the data scientist's data set was blindly annotated by the health care specialist, allowing to calculate the kappa coefficient. The kappa coefficient for interrater reliability for the presence of a general HRQoL impact was 0.724; for the physical impact, it was 0.871; for the psychic impact, 0.663; for the activity-related impact, 0.639; and for the relational impact, 0.649; this is while no messages mentioned a financial impact in this subset (κ=0). Thus, agreement ranged from strong to very strong according to the kappa Cohen coefficient scale for 4 of the dimensions, but not for the financial one because no financial impact was mentioned in the subset of messages. This high interrater reliability for 4 of the 5 dimensions suggests that the used guidelines and training about the HRQoL ensured a homogeneous annotation of the messages.

Preprocessing and Feature Engineering

All impact-related messages were used to generate dimension-specific features. Other features were based on the message structure, such as expressed sentiment (eg, positive, negative, anger, disgust, fear, joy, sadness, and surprise), grammar (eg, count of pronouns, who is writing, and negative sentences), and conjugation (eg, count of verb tenses). A lexical field score corresponding to each HRQoL dimension was computed by counting the associated expressions previously collected during the annotation stage. We used the R packages of the Detec't extractor [39,40] to create lexical variables. This phase enabled the development of specific models of impact detection per dimension. The rationale behind this process was to be able to adapt to the many expressions of the patients. Psychic impacts and physical impacts are different, and so are the expressions used to describe them. Hence, having specific models by dimension is a way to minimize an interpretation bias.

We ended the process with a data set or corpus composed of quantitative features such as expressed sentiments (from the Linguistic Inquiry and Word Count dictionary), grammar, conjugation, and lexical fields of HRQoL-related features.

Model Selection

We used data mining and machine learning technologies to categorize and analyze retrieved data of our final corpus

according to our predefined objective. As our features do not exhibit negative values, we normalized our data by dividing all feature values by their respective maximum so that all values would be somewhere between 0 and 1, thus minimizing interclass and intraclass variances. All the missing values were replaced by the median so as not to influence intraclass variance.

We obtained a first classification algorithm to determine if there was an impact on HRQoL (corresponding to the first step of manual annotation). Subsequently, we created a classification algorithm for each dimension to assess whether the impact concerned the related dimension (second step).

We used a 5-fold sequential forward floating selector with an extreme gradient boosting algorithm to select the best features combination. We tried first to maximize the model accuracy, but we ended up with several false negative cases. We finally chose the area under the curve (AUC) as our scoring method to maximize the true positive rate because we would rather have a slightly larger number of posts containing an impact, even with false positive, than missing some of these.

We chose sequential forward floating selector over LASSO (least absolute shrinkage and selection operator) to maximize the ROC (receiver operating characteristic) value, while LASSO is trying to minimize the cost function. This allowed to obtain the best performances for all classes instead of the majority class.

We then tried several machine learning algorithms, the K-Nearest Neighbors, SVM, Multi-Layer Perceptron, Random Forest, and finally XGBoost.

Except for the psychic dimension, XGBoost was far above the other methods in terms of AUC (Table 1).

We then performed a 5-fold cross-validated grid search on our selected features to tune our hyperparameters. We split our training set into 5 samples and trained the algorithm successively on 4 of these samples, while the last sample was used as validation set. This method allowed minimizing overfitting and making sure that the models generalize well. We varied the learning rate, the number of epochs, the number of trees and their maximum depth, the minimum weight needed in a child node, the minimum loss reduction required to make a further partition on a leaf node, and the L1 regularization. LASSO regression was preferable for feature selection in case of a great number of features, making nonimportant features even more insignificant in term of weights.

Table 1. AUC (area under the curve) values of the different machine learning methods.

Algorithm	Impact[a]	Physical	Psychic	Activity	Relational	Financial
KNN[b]	69.9	66.5	64.9	64.4	68.6	65.6
SVM[c]	67.9	63.3	56.3	55.3	57.7	56.9
MLP[d]	74.6	67.9	61.6	64.5	64.5	58.6
RF[e]	75	70.5	70.7	71.8	70.9	69
XGB[f]	78.5	75	71	76	71.7	76.5

[a]At least 1 impact on at least 1 of the 5 dimensions.

[b]KNN: K-Nearest Neighbors.

[c]SVM: support vector machine.

[d]MLP: Multi-Layer Perceptron.

[e]RF: Random Forest.

[f]XGB: XGBoost.

This process allowed elaborating a model that can detect a general impact. The developed algorithm filtered the corpus of messages into 2 categories: HRQoL impacted or not. For each model, we selected the relevant variables by applying the sequential forward floating selector and chose which combination could better separate an impact message from a nonimpact message. In a nutshell, it removes or adds one feature at a time on the classifier and test performances until it reaches the best possible score. The same steps were then reproduced in each dimension according to their specific features in order to obtain specific algorithms fitted for each dimension.

Features of patient expressions specific to each impact were identified with the Linguistic Inquiry and Word Count dictionary, which provides expressions for various feelings, such as positivity, negativity, joy, sadness, disgust, surprise, fear, and anger. The frequency of these expressions within the posts was used to select the relevant variables for each impact domain (Table 2). Patterns identified during data labelling were also used to select relevant variables. We can assume than to describe daily actions and difficulties, the present tense is the most appropriate tense. Conversely, to talk about an impact within the family, “we” is more often used.

Due to the lack of a specific dimension’s impact mention, some classes were imbalanced regarding one another; in order to correct that, we created an artificially balanced class by using the oversampling method SMOTE (synthetic minority oversampling technique) [41]. Based on the mathematical structure of the under-represented messages, this technique artificially creates similar examples that fit the same feature pattern in order to balance the categories. We used this method for the activity-related, relational, and financial impact algorithms.

Table 2. Most important features by model.

Dimension	Feature 1	Feature 2	Feature 3
Impact	Number of infinitive verbs	Count of first person of singular markers	Counted sadness expressions
Physical	Counted physical related expressions	Counted negative expressions	Number of negations
Psychic	Counted psychic-related expressions	Counted anger expressions	Counted fear expressions
Activity-Related	Counted professional, academic, or daily activity-related expressions	Count of verbs in present tense	Number of pronouns
Relational	Counted relational expressions	Count of first person of plural markers	Count of past participle verbs
Financial	Count of financial expressions	Number of "not"	Count of verbs in past tense

Statistical Analysis

We used sensitivity (defined as correctly identifying an HRQoL impact when classified as so by our algorithm) and specificity (defined as correctly identifying a message without impact when classified as so by our algorithm). The ROC curve and the AUC were considered to measure the overall performance of the algorithm. The ROC curve represented the true positive rate (sensitivity) plotted in function of the false positive rate (100-specificity) for different thresholds of the metric.

Results

Corpus

We extracted 20,000 messages from health-related forums mentioning diverse and different diseases such as cancers, diabetes, endometriosis, and psychological afflictions, from defined diagnosis to syndrome name (eg, nausea, "feeling blue/depressed"). Treatments such as vaccines, Levothyrox (thyroid hormones) and psychiatric drugs were also mentioned. The goal was to constitute a representative panel of health impairments, including physical, psychological, frequent, rare, light, and heavy afflictions. This corpus merged random messages mentioning 1280 medical terms (at least 1 term per message, disease, or medication). The diseases and treatment terms were identified with exact matching methods on MedDRA (Medical Dictionary for Regulatory Activities). Of the 20,000 extracted messages posted from 2000 to 2019, we randomly selected 3000 (15%) messages, which were split into 1000 and 2000. We removed duplicate entries so that we finally annotated 1399 messages: 1000 (71%) related to diseases and 399 (29%) to treatments. In the end, we had 818 (58%) messages showing at least 1 impact on QoL, 442 (31%) showing physical impact, 519 (37%) psychic, 363 (25%) activity-related, 193 (13%) relational, and 69 (4%) financial (Table 3). Many impacts on more than 1 dimension can be expressed in messages by patients.

The final corpus was then composed of 1399 French forum messages extracted from 19 conversation threads. These messages were written by users in an informal style. The length ranged from a few words to narratives longer than 1000 characters, the average message length being 905 (SD 1041) characters.

Table 3. Number of messages showing health-related quality of life impact, at least 1 impact, and by dimension.

Dimension	Message, n (%)
At least 1 impact	818 (58)
Physical	442 (31)
Psychic	519 (37)
Activity-Related	363 (25)
Relational	193 (13)
Financial	69 (4)

Modeling

From our 1399 annotated messages, we chose to split them in a 70:30 ratio where 70% of the messages were used for the training phase and the rest as validation. Out of the 1399 messages, 420 (30%) were used to evaluate the model. Among these 420 messages, 203 (48%) were predicted with an impact.

We searched for lexical fields in order to evaluate the attribution of a score per dimension. We tested the different machine learning algorithms to optimize the parameters and the results. Extreme gradient boosting was the chosen model for both impact detection and specific dimension identification. The final HRQoL impact detection algorithm was composed of several models, including a model that identified the presence of an impact and all the impact-flagged messages, which went through each specific dimension model. The models were trained using cross-validation and hyperparameter optimization. Oversampling was used to augment infrequent dimensions. This allowed us to detect a general impact on HRQoL with a sensitivity of 0.8 and a specificity of 0.7 (Table 4). Overall, 818 messages presented an impact and 581 did not. For physical impact, sensitivity was 0.56, and specificity was 0.857; for psychic impact, 0.58 and 0.828; for activity-related impact, 0.71 and 0.79; for relational impact, 0.675 and 0.73; and for financial impact, 0.77 and 0.814, respectively.

Table 4. Overall results of the different HRQoL[a]-impacted dimensions.

Dimension	F-measure	ROC[b] curve
At least 1 impact	78.6	78.5
Physical	70	75
Psychic	68	71
Activity-Related	75	76
Relational	70.6	71.7
Financial	76	76.5

[a]HRQoL: health-related quality of life.

[b]ROC: receiver operating characteristic.

Discussion

Principal Findings

We developed an algorithm to evaluate the impact diseases and treatments can have on patients' HRQoL based on their emotions and opinions shared on social media. The algorithm was based on an adaptation for the social media listening approach, of the EQ-5D and SF-36 scales, which are recommended by several national and international institutions for assessing HRQoL and whose psychometric proprieties are well known [7,42,43]. Five dimensions of impact on HRQoL were then covered and identified in a filtered corpus of 1399 messages. The algorithm was able to detect different types of disease and treatment impact on HRQoL with good sensitivity and specificity. The algorithm had an ROC score of 0.785 for detecting at least 1 impact on at least 1 of the 5 dimensions (0.75 for physical dimension, 0.71 for psychic, 0.76 for activity-related, 0.717 for relational, and 0.765 for financial). Compared to other studies [44,45], these indicators were high and robust; for example, with Twitter and Facebook data, the area under the curve of Caster et al [44] varied between 0.43 and 0.67. For patient forum posts, sensitivity was 0.14 (and specificity was 0.88); and for Twitter and Facebook, sensitivity was 0.08 or lower. However, the objectives and approaches of these studies were different from ours, and it is thus quite difficult to compare the results. Performance might vary according to the data source. Considering that we were able to access a large data set and to use a satisfying training subset, this might explain our better performance. Nonetheless, Facebook and Twitter were discarded from our extracted sources due to the short messages of Twitter and the difficulties of access to Facebook data.

Social media listening allows direct monitoring of patients' messages capturing "live" their opinions and feelings compared to a punctual "fixed" self-administered questionnaire. This approach corresponds more to the evolutive nature of HRQoL.

Our study adds to the literature on the use of NLP and text mining concerning medical care from web-based data. This approach relies on the potential strength of large and real time web-based data, which are complementary to classic medical reporting systems. This work contributes to the need for an improvement in methodologies that can produce more sophisticated joint models of user and message-level information or the use of syntactic structure as their features.

A similar study was conducted to outpredict baselines of popular happy and hedonistic lexica through the satisfaction with life scale over Facebook volunteers [46]. The findings of this study were also encouraging by demonstrating the effectiveness of machine learning algorithms to detect users' health-related emotions.

Another study carried out in France [47] showed a good performance in terms of sensitivity and specificity of an NLP method to detect self-reported signals of issues with treatments. Our results confirm the same success of established statistical detection algorithms in social media for a wide range of diseases and treatments.

Strengths and Limitations

This methodological study contributes to the growing research on social media listening and machine learning in general as a technique to develop and train tools to measure broad constructs such as HRQoL. Our work is among the first research projects proving that a social media listening tool can provide a sound and efficient measurement of impacts on HRQoL directly accessible from patients to health professionals. In this sense, it highlights some of the promises of social media and forums as data sources. One of the strengths of this study was the quality of preprocessing and processing of the data extracted. Several cleansing and validation steps were performed to ensure the quality of the messages. Furthermore, we used medically validated (general) scales, the EQ-5D and SF-36, as a strong scientific basis and gold standard for the detection of 5 specific HRQoL dimensions (ie, physical, psychological, activity-related, financial, and social). Different diseases or treatments would differently affect patients; therefore, our generalist approach of the machine learning model, which has been trained based on the patients' free speech on various diseases and treatments, is able to detect different expressions of impact on our 5 common dimensions.

However, an algorithm does not have the human sensitivity to understand very specific and subjective ways of expressing a HRQoL impact (such as sarcasm), despite the constant improvement of the work. Sentimental analysis can complement such algorithms, and manual review remains strongly required. Additionally, our approach lacks flexibility in the feature extraction process; impact-specific features are not exhaustive because the expression of impact can vary. This also requires improvement in order to complete the lexical fields.

Limitations also include the data sources. More analysis is needed to prove that insights from social media are complementary to a patient-centric repository. Furthermore, Twitter and Facebook were discarded as sources due to short message format and accessibility issues; however, this does not mean that these social platforms are irrelevant resources for analyzing health testimonies from patients.

Our data were randomly extracted from a large sample of French messages coming from French forums and social media. The fact that our sample selection was random should ensure a certain representativity of the internet message population. The proportion of women speaking about their health in forums is higher than the proportion of men (difference of 6%) [48], which introduces a possible bias when exploring HRQoL. However, our algorithm is designed to work on data coming from French forums and social media with similar gender proportions.

Future work is needed to continue training the algorithm and to further study the differences on HRQoL between internet users and patients not posting messages on social media or forums.

Implications

We provided evidence that social media listening can be used to assess the impact and burden of one or more diseases and treatments on patients' HRQoL. These findings can provide public health experts, health care professionals, and pharmaceutical companies with patient-generated information on their experiences with treatments, burden of diseases, and needs for appropriate medical care in a timely manner and in real-life conditions. For instance, the generated data coming directly from patients can inform potential changes of a treatment and development of new pharmaceutical products. The use of social media listening might be recommended to monitor HRQoL constantly and consistently in patients under a new treatment or experiencing a severe disease.

Conclusion

We developed an algorithm that can translate social media patient messages into the identification of an impact on HRQoL. Based on medically validated questionnaires, this is a patient-centered approach using machine learning and NLP to better understand how diseases and treatments can represent a burden for patients.

Acknowledgments

None declared.

Conflicts of Interest

SR, TM, MK, PF, PV, AM, NT, and SS are members of Kap Code. IM is a member of the Bordeaux Population Health Research Center, UMR 1219, Bordeaux University, Inserm.

Multimedia Appendix 1
Literature references.
[DOCX File , 24 KB - jmir_v24i1e31528_app1.docx]

References

1. Cline RJ, Haynes KM. Consumer health information seeking on the Internet: the state of the art. Health Educ Res 2001 Dec;16(6):671-692. [doi: 10.1093/her/16.6.671] [Medline: 11780707]
2. Morlane-Hondère F, Grouin C, Zweigenbaum P. Identification of Drug-Related Medical Conditions in Social Media. 2016 May Presented at: Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC; 2016; Portoroz, Slovenia.
3. Bansal G, Zahedi F, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. Decision Support Systems 2010 May;49(2):138-150. [doi: 10.1016/j.dss.2010.01.010]
4. Nussbaum M, Sen A. The Quality of Life. Oxford, UK: Clarendon Press; 1993.
5. Measuring Quality of Life.: World Health Organization; 1997. URL: https://tinyurl.com/bvf746fa [accessed 2022-01-21]
6. Karimi M, Brazier J. Health, Health-Related Quality of Life, and Quality of Life: What is the Difference? Pharmacoeconomics 2016 Jul 18;34(7):645-649. [doi: 10.1007/s40273-016-0389-9] [Medline: 26892973]
7. Hays RD, Reeve BB. Epidemiology and Demography in Public Health. In: Measurement and Modeling of Health-Related Quality of Life. San Diego, USA: International Encyclopedia of Public Health; 2010.
8. Koonal KS. A brief review of concepts: health, quality of life, health-related quality of life and well being. Rotterdam, the Netherlands: EuroQol Research Foundation; 2017:9.
9. Lorente S, Vives J, Viladrich C, Losilla J. Tools to assess the measurement properties of quality of life instruments: a meta-review protocol. BMJ Open 2018 Jul 23;8(7):e022829 [FREE Full text] [doi: 10.1136/bmjopen-2018-022829] [Medline: 30037880]
10. EuroQol Group. EuroQol - a new facility for the measurement of health-related quality of life. Health Policy 1990 Dec;16(3):199-208. [doi: 10.1016/0168-8510(90)90421-9]

11. Ware J, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey Manual Interpretation Guide. Boston, USA: The Health Institute, New England Medical Center; 1993.
12. Quality of Life Research Group. Pushing the boundaries: frontiers of quality of life research. Qual Life Res 2012 Jan;20 Suppl 1:2-117. [doi: 10.1007/s11136-011-0097-z] [Medline: 22298117]
13. Cotté FE, Voillot P, Bennett B, Falissard B, Tzourio C, Foulquié P, et al. Exploring the Health-Related Quality of Life of Patients Treated With Immune Checkpoint Inhibitors: Social Media Study. J Med Internet Res 2020 Sep 11;22(9):e19694 [FREE Full text] [doi: 10.2196/19694] [Medline: 32915159]
14. Powell GE, Seifert HA, Reblin T, Burstein PJ, Blowers J, Menius JA, et al. Social Media Listening for Routine Post-Marketing Safety Surveillance. Drug Saf 2016 May 21;39(5):443-454. [doi: 10.1007/s40264-015-0385-6] [Medline: 26798054]
15. Dreisbach C, Koleck TA, Bourne PE, Bakken S. A systematic review of natural language processing and text mining of symptoms from electronic patient-authored text data. Int J Med Inform 2019 May;125:37-46 [FREE Full text] [doi: 10.1016/j.ijmedinf.2019.02.008] [Medline: 30914179]
16. New update of official guide to the methods of economic evaluation for France recommends use of EQ-5D-5L to derive a utility score. EQ-5D. 2020 Sep. URL: https://tinyurl.com/bdd2vwz5 [accessed 2022-01-21]
17. Haute Autorité de Santé. Choix méthodologiques pour l'évaluation économique à la HAS. Saint-Denis La Plaine, France: HAS; 2020:8.
18. Young IJB, Luz S, Lone N. A systematic review of natural language processing for classification tasks in the field of incident reporting and adverse event analysis. Int J Med Inform 2019 Dec;132:103971. [doi: 10.1016/j.ijmedinf.2019.103971] [Medline: 31630063]
19. Forums. Atoute. URL: https://forum.atoute.org/ [accessed 2022-01-21]
20. Forums Santé. Doctissimo. URL: https://forum.doctissimo.fr/ [accessed 2022-01-21]
21. Forum. aufeminin. URL: https://forum.aufeminin.com/forum/ [accessed 2022-01-21]
22. Forum Journal des Femmes Santé. Journal des Femmes Santé. URL: https://sante-medecine.journaldesfemmes.fr/forum/ [accessed 2022-01-21]
23. Psychoactif, l'espace solidaire entre usagers de drogues. Psychoactif. URL: https://www.psychoactif.org/ [accessed 2022-01-21]
24. HardWare.fr. URL: https://forum.hardware.fr/ [accessed 2022-01-21]
25. Le forum cancer du sein des Impatientes. Les Impatientes. URL: http://www.lesimpatientes.com/ [accessed 2022-01-21]
26. Laxophobie et Colopathie fonctionnelle. Laxophobie. URL: http://www.laxophobie.fr/ [accessed 2022-01-21]
27. Forum. magicmaman. URL: http://forum.magicmaman.com/ [accessed 2022-01-21]
28. Forum de discussion. Vivre sans thyroïde. URL: https://www.forum-thyroide.net/ [accessed 2022-01-13]
29. Forum Ados. Public.fr. URL: https://www.public.fr/Forum-Ados [accessed 2022-01-21]
30. Perles des forums. Onmeda. URL: https://www.onmeda.fr/ [accessed 2022-01-21]
31. Les Forums. Psychologies.com. URL: https://forum.psychologies.com/ [accessed 2022-01-21]
32. Expériences avec des médicaments. meamedica. URL: https://www.meamedica.fr/ [accessed 2022-01-21]
33. Forum FS Generation. Futura. URL: https://forums.futura-sciences.com/ [accessed 2022-01-21]
34. AlloDocteurs. URL: https://www.allodocteurs.fr/ [accessed 2022-01-21]
35. Forum Santé. Vulgaris Médical. URL: https://www.vulgaris-medical.com/forum-sante [accessed 2022-01-21]
36. Informer et soutenir Lutter contre le lymphome. France Lymphome Espoir. URL: https://www.francelymphomeespoir.fr/ [accessed 2022-01-21]
37. Forum de discussion. Maman pour la vie. URL: https://www.mamanpourlavie.com/forum/ [accessed 2022-01-21]
38. Cro spécialiste en real world evidence. Kappa Santé. URL: https://www.kappasante.com/ [accessed 2022-01-21]
39. Abdellaoui R, Foulquié P, Texier N, Faviez C, Burgun A, Schück S. Detection of Cases of Noncompliance to Drug Treatment in Patient Forum Posts: Topic Model Approach. J Med Internet Res 2018 Mar 14;20(3):e85 [FREE Full text] [doi: 10.2196/jmir.9222] [Medline: 29540337]
40. Abdellaoui R, Schück S, Texier N, Burgun A. Filtering Entities to Optimize Identification of Adverse Drug Reaction From Social Media: How Can the Number of Words Between Entities in the Messages Help? JMIR Public Health Surveill 2017 Jun 22;3(2):e36 [FREE Full text] [doi: 10.2196/publichealth.6577] [Medline: 28642212]
41. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. jair 2002 Jun 01;16:321-357. [doi: 10.1613/jair.953]
42. Chen T, Li L, Kochen MM. A systematic review: How to choose appropriate health-related quality of life (HRQOL) measures in routine general practice? J. Zhejiang Univ. Sci 2005;6B(9):936-940. [doi: 10.1631/jzus.2005.b0936]
43. Lins L, Carvalho FM. SF-36 total score as a single measure of health-related quality of life: Scoping review. SAGE Open Med 2016 Oct 04;4:2050312116671725 [FREE Full text] [doi: 10.1177/2050312116671725] [Medline: 27757230]
44. Caster O, Dietrich J, Kürzinger ML, Lerch M, Maskell S, Norén GN, et al. Assessment of the Utility of Social Media for Broad-Ranging Statistical Signal Detection in Pharmacovigilance: Results from the WEB-RADR Project. Drug Saf 2018 Dec;41(12):1355-1369 [FREE Full text] [doi: 10.1007/s40264-018-0699-2] [Medline: 30043385]

45. Colilla S, Tov EY, Zhang L, Kurzinger M, Tcherny-Lessenot S, Penfornis C, et al. Validation of New Signal Detection Methods for Web Query Log Data Compared to Signal Detection Algorithms Used With FAERS. Drug Saf 2017 May 2;40(5):399-408. [doi: 10.1007/s40264-017-0507-4] [Medline: 28155198]
46. Schwartz HA, Sap M, Kern M, Eichstaedt JC, Kapelner A, Agrawal M, et al. Predicting individual well-being through the language of social media. Pac Symp Biocomput 2016;21:516-527 [FREE Full text] [Medline: 26776214]
47. Kürzinger ML, Schück S, Texier N, Abdellaoui R, Faviez C, Pouget J, et al. Web-Based Signal Detection Using Medical Forums Data in France: Comparative Analysis. J Med Internet Res 2018 Nov 20;20(11):e10466 [FREE Full text] [doi: 10.2196/10466] [Medline: 30459145]
48. Parler de sa santé en ligne : une pratique loin d'être marginale et qui peut aider la recherche. Odoxa. URL: http://www.odoxa.fr/sondage/parler-de-sante-ligne-pratique-loin-detre-marginale-aider-recherche/ [accessed 2022-01-21]

Abbreviations

AUC: area under the curve
EQ-5D: Euro Quality of Life 5 Dimensions
FACT-G: Functional Assessment of Cancer Therapy—General
GBDT: gradient boosting decision tree
HRQoL: health-related quality of life
LASSO: least absolute shrinkage and selection operator
MedDRA: Medical Dictionary for Regulatory Activities
NER: named entity recognition
NLP: natural language processing
POS: part of speech
QLQ-C30: Quality of Life Questionnaire
QoL: quality of life
ROC: receiver operating characteristic
SF-36: Short Form Health Survey
SMOTE: synthetic minority oversampling technique
SVM-L: support vector machine light
SVM-R: support vector machine regression
TF-IDF: term frequency-inverse document frequency

Edited by R Kukafka; submitted 24.06.21; peer-reviewed by A Trifan, D Huang; comments to author 11.08.21; revised version received 05.10.21; accepted 29.10.21; published 28.01.22.

Please cite as:
Renner S, Marty T, Khadhar M, Foulquié P, Voillot P, Mebarki A, Montagni I, Texier N, Schück S
A New Method to Extract Health-Related Quality of Life Data From Social Media Testimonies: Algorithm Development and Validation
J Med Internet Res 2022;24(1):e31528
URL: https://www.jmir.org/2022/1/e31528
doi:10.2196/31528
PMID:35089152

Original Paper

Characterizing Patient-Clinician Communication in Secure Medical Messages: Retrospective Study

Ming Huang[1], PhD; Jungwei Fan[1,2], PhD; Julie Prigge[3], MA; Nilay D Shah[2,4], PhD; Brian A Costello[3], MD; Lixia Yao[1], PhD

[1]Department of Artificial Intelligence and Informatics, Mayo Clinic, Rochester, MN, United States

[2]Center for the Science of Health Care Delivery, Mayo Clinic, Rochester, MN, United States

[3]Center for Connected Care, Mayo Clinic, Rochester, MN, United States

[4]Department of Quantitative Health Sciences, Mayo Clinic, Rochester, MN, United States

Corresponding Author:
Lixia Yao, PhD
Department of Artificial Intelligence and Informatics
Mayo Clinic
200 First Street SW
Rochester, MN, 55905
United States
Phone: 1 507 293 7953
Fax: 1 507 284 1516
Email: lixia.cn.yao@gmail.com

Abstract

Background: Patient-clinician secure messaging is an important function in patient portals and enables patients and clinicians to communicate on a wide spectrum of issues in a timely manner. With its growing adoption and patient engagement, it is time to comprehensively study the secure messages and user behaviors in order to improve patient-centered care.

Objective: The aim of this paper was to analyze the secure messages sent by patients and clinicians in a large multispecialty health system at Mayo Clinic, Rochester.

Methods: We performed message-based, sender-based, and thread-based analyses of more than 5 million secure messages between 2010 and 2017. We summarized the message volumes, patient and clinician population sizes, message counts per patient or clinician, as well as the trends of message volumes and user counts over the years. In addition, we calculated the time distribution of clinician-sent messages to understand their workloads at different times of a day. We also analyzed the time delay in clinician responses to patient messages to assess their communication efficiency and the back-and-forth rounds to estimate the communication complexity.

Results: During 2010-2017, the patient portal at Mayo Clinic, Rochester experienced a significant growth in terms of the count of patient users and the total number of secure messages sent by patients and clinicians. Three clinician categories, namely "physician—primary care," "registered nurse—specialty," and "physician—specialty," bore the majority of message volume increase. The patient portal also demonstrated growing trends in message counts per patient and clinician. The "nurse practitioner or physician assistant—primary care" and "physician—primary care" categories had the heaviest per-clinician workload each year. Most messages by the clinicians were sent from 7 AM to 5 PM during a day. Yet, between 5 PM and 7 PM, the physicians sent 7.0% (95,785/1,377,006) of their daily messages, and the nurse practitioner or physician assistant sent 5.4% (22,121/408,526) of their daily messages. The clinicians replied to 72.2% (1,272,069/1,761,739) patient messages within 1 day and 90.6% (1,595,702/1,761,739) within 3 days. In 95.1% (1,499,316/1,576,205) of the message threads, the patients communicated with their clinicians back and forth for no more than 4 rounds.

Conclusions: Our study found steady increases in patient adoption of the secure messaging system and the average workload per clinician over 8 years. However, most clinicians responded timely to meet the patients' needs. Our study also revealed differential patient-clinician communication patterns across different practice roles and care settings. These findings suggest opportunities for care teams to optimize messaging tasks and to balance the workload for optimal efficiency.

(J Med Internet Res 2022;24(1):e17273) doi:10.2196/17273

KEYWORDS

patient portal; secure message; patient-clinician communication; workload; response time; message round

Introduction

A patient portal is a secure online platform that allows patients to conveniently access and manage personal health information and communicate with their clinicians [1]. After the Health Information Technology for Economic and Clinical Health Act of 2009, patient portals have gained widespread adoption by health care systems in the United States [2,3]. In 2017, over 90% of health care organizations including the Veterans Administration, Mass General Brigham, Kaiser Permanente, and Mayo Clinic offered patient portal access to their patients [4]. Patient portals give patients 24-7 access to their health information (eg, clinical visits, lab test results, medications, and discharge summaries) from anywhere with internet connection [5] and have been shown to improve patient self-management by promoting the awareness of disease knowledge, status, and progress [2].

A significant function in patient portals is patient-clinician secure messaging, which enables patients and clinicians to timely communicate on a wide spectrum of issues. Patients use secure messaging to request medical appointments and refill prescriptions online [6,7]. Clinicians send patients appointment reminders and promote timely preventative care [8,9]. Patients and clinicians can communicate back and forth on complex situations such as new symptoms, disease follow-ups, medication concerns, and other medical questions. Evidence suggests that secure messaging improves health care efficiency, productivity, and quality. For instance, Zhou et al [10] investigated more than 4000 users at Kaiser Permanente before and after the introduction of a secure message system and found that their annual rates of in-person primary care visits were reduced by 9.7% after their adoption of secure messaging. Simon et al [11] showed that both antidepressant adherence rate and depression treatment satisfaction increased by 20% among patients using secure messaging.

On the other hand, the secure messaging resulted in additional workload for clinicians and contributed to work burnout, as it increased patient-clinician interactions between in-person patient visits. According to a survey, 63% of 43 clinicians across 5 clinics disagreed with the notion that “secure messaging reduces my workload,” and 33% agreed with the notion that “secure messaging has a negative effect on my workflow” [12]. Another study shows that primary care physicians spend, on average, 1.4 hours of their workday (5.9 hours) interacting with electronic health records for non–face-to-face care after clinic hours [13].

With the increase in the number of patients signing up for these portals, the number of secure messages has risen substantially, especially during the COVID-19 pandemic [14-18]. It will be critical to the care teams to understand the patient-clinician messaging and to properly distribute the communication load for better efficiency and avoiding clinician burnout. It could be foreseen that some health care systems would be likely to face the challenge of managing the increasing volume of patient messages soon [12], which will require new billing models and practice metrics, or additional infrastructures, including support staffs to reply to the increasing volume of patient messages. However, there is limited understanding of the use of secure messaging and the extent of users’ interaction through this medium including clinician messaging load, messaging time delay, messaging time distribution in a day, and messaging complexity of a communication thread.

In this study, we attempted to bridge this knowledge gap by analyzing more than 5 million secure messages that were generated by patients and clinicians between 2010 and 2017 at a large multispecialty health system at Mayo Clinic, Rochester. We performed message-oriented and sender-oriented analyses by calculating the message volumes, patient or clinician population sizes, message counts per patient or clinician, time distribution of clinician messaging, and their trends over the years. We also performed thread-oriented analysis to probe the time delay in clinician responses to patient messages to assess their communication efficiency and the back-and-forth rounds to estimate the communication complexity. Our findings shed light on the patient-clinician digital communication and inform future improvement in the use of secure medical messaging.

Methods

Data Collection and Preprocessing

The patient portal (Patient Online Services) at Mayo Clinic, Rochester [19] was started in 2010 for primary care practice and later extended to specialty practice in 2013. The patient portal allows patients and clinicians to communicate bidirectionally via secure messaging on a wide range of issues. We retrieved more than 5 million secure messages from the patient portal between February 18, 2010, and December 31, 2017. Each message has a unique identifier (ID), previous message ID, initial message ID, sender ID, recipient ID, the timestamp when it was sent, message subject, and message body. In a message thread with a series of back-and-forth messages, the initial message is the first message initiated by a sender, and the message ID of the initial message serves as the ID of the message thread. We then applied three filters: exclude the messages with empty message bodies; exclude the messages sent by mock-up patients and clinicians that were created for testing; and exclude messages sent by a clinician group where the sender uses a shared ID, usually for impersonal communication. In the end, we obtained a total of 5,654,514 secure messages sent by both patients and clinicians for the following analysis.

Message-Oriented and Sender-Oriented Analysis

We started by calculating the descriptive statistics for the approximately 5.6 million secure messages to probe four aspects of patient-clinician communication. (1) The total numbers (volumes) of messages sent by patients and clinicians and the overall counts of unique patient and clinician senders. We also distinguish whether a message is an initiated message from a sender (ie, a drug-related question from a patient or an

appointment reminder from a clinician) or a replied message (ie, a follow-up question or clarification) in a message thread; (2) For patient-sent messages, we calculated the distribution of message counts per patient for the entire study period, the number of secure messages and the count of unique patient senders by year, and the distribution of message counts per patient each year. The patients who registered for patient portal but did not send any messages were excluded; (3) Similarly, for clinician-sent messages, we calculated the distribution of message counts per clinician for the entire study period, the number of secure messages and the count of clinician senders by year, and the distribution of message counts per clinician each year. In addition, we grouped the clinicians into 9 categories based on their practice roles (ie, physician, nurse practitioner/physician assistant [NP/PA], registered nurse [RN], and other) and care settings (ie, primary care, specialty, and other) as listed in Table 1, and measured the workload for each clinician category. The "Other—other" category refers to other supporting staffs who communicated with patients via secure messaging, such as patient appointment service specialists, social workers, and financial counselors who work outside of the primary and specialty care setting; and (4) To analyze the workload of clinicians in different times of a day, we split 24 hours into 12 time slices for each day (ie, 11 PM to 1 AM, 1 to 3 AM, 3 to 5 AM, 5 to 7 AM, 7 to 9 AM, 9 to 11 AM, 11 AM to 1 PM, 1 to 3 PM, 3 to 5 PM, 5 to 7 PM, 7 to 9 PM, and 9 to 11 PM) and calculated the percentage of secure messages that those clinicians sent in each of the 12 time slices by year.

Table 1. Clinician categories based on their practice roles and care settings.

Role	Primary care	Specialty	Other
Physician	Physician—primary care	Physician—specialty	N/A[a]
NP[b]/PA[c]	NP/PA—primary care	NP/PA—specialty	N/A
RN[d]	RN—primary care	RN—specialty	N/A
Other	Other—primary care	Other—specialty	Other—other

[a]N/A: not applicable.

[b]NP: nurse practitioner.

[c]PA: physician assistant.

[d]RN: registered nurse.

Thread-Oriented Analysis

We investigated two aspects of patient-clinician communication within message threads. The first aspect is the time delay of clinician responses to patient messages. The analysis of time delay between patient messages and clinician responses may suggest how promptly clinicians responded to patients. We identified all the pairs of patient-sent messages and clinician-replied messages in a message thread and calculated the time difference (in days) between them. We then calculated the distribution of time delays for the entire study period and for each year. Secondly, patients often communicate with clinicians back and forth for multiple times in a message thread. We examined these message threads by measuring the number of back-and-forth rounds in each message thread (ie, length of a message thread) and calculating the distribution of message threads in terms of message thread length over time.

We developed a series of scripts in Python (Python Software Foundation) together with popular Python libraries (eg, pandas [20], NumPy [21], SciPy [22], and Matplotlib [23]) to perform data collection and preprocessing, statistical analysis, and visualization. No patients were exposed to any intervention. We used the data from the Mayo Clinic Unified Data Platform for analysis. The study was approved by the Mayo Clinic Institutional Review Board (19-002211).

Results

Message-Oriented and Sender-Oriented Analysis

Descriptive Statistics of Secure Medical Messages and Their Senders

Table 2 lists the total numbers of secure messages sent (ie, initiated and replied) by patients and clinicians as well as unique patient senders and clinician senders. The number of messages initiated by clinicians was almost identical to that of messages replied by clinicians but was slightly more than that of messages initiated by patients (1.7 million versus 1.5 million). However, the number of patient-initiated messages was about 3 times that of patient-replied messages (1.5 million versus 0.5 million). In addition, 93.7% of the patients (203,166/216,740) had initiated secure messages, whereas only 52.6% (113,974/216,740) had replied to their clinicians. This suggests that the patients in general initiated a message when they had health-related issues, rather than replying in a message thread, whereas the clinicians were obligated to initiate messages and respond to patient messages.

Table 2. Descriptive statistics of secure medical messages and their senders for the 8-year period.

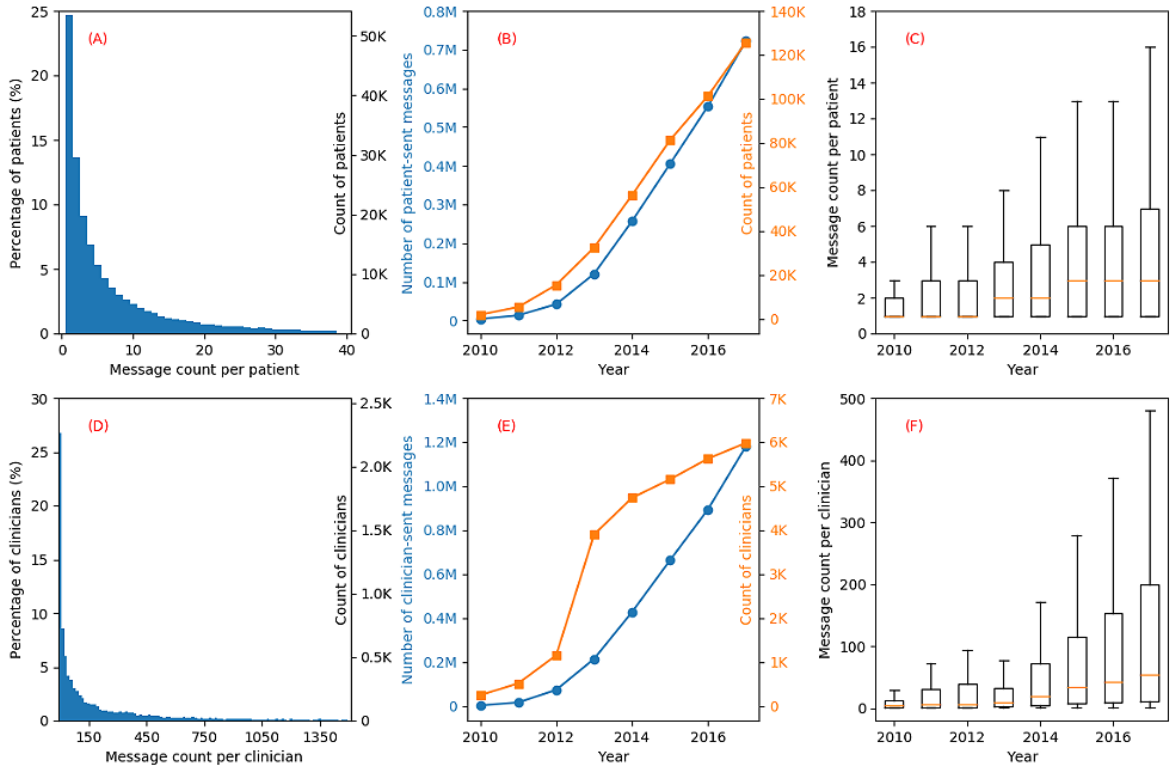
Senders and secure message types	Numbers of secure messages, n (%)	Counts of unique senders, n (%)
Patient		
Initiated messages	1,569,172 (74.06)	203,166 (93.74)
Replied messages	549,601 (25.94)	113,974 (52.59)
Total	2,118,773	216,740
Clinician		
Initiated messages	1,774,000 (50.17)	5690 (67.27)
Replied messages	1,761,741 (49.83)	8,070 (95.40)
Total	3,535,741	8459

The ratio of overall unique patient count to overall unique clinician count was about 25:1 (216,740 versus 8450). The ratio of patient count and clinician count per year increased from 7.7 (1919 versus 249) in 2010 to 21.0 (125,647 versus 5980) in 2017 as shown in Table S1 (Multimedia Appendix 1). This indicates that the workload per clinician had an increasing trend over years in terms of patient users on the secure messaging system.

Patient-Sent Secure Messages

We illustrated the distribution of message count per patient, the total numbers of patient-sent messages and unique patients by year, and the box-whisker graph of message count per patient by year in Figure 1 (A-C). We found that 95.4% (206,818/216,740) of the patients sent less than 40 messages, as shown in Figure 1 (A). The median number of patient-sent messages was 4. The maximum number of patient-initiated and patient-replied messages was 2052 and 302, respectively (Table S1, Multimedia Appendix 1). We observed a similar increasing trend between the numbers of unique patients and patient-sent messages over years, as shown in Figure 1 (B), indicating the strong adoption of this technology. The Pearson correlation coefficient between the numbers of unique patients and patient-sent messages calculated with SciPy is r=1. The median of message count per patient increased from 1 to 3 during 2010-2017 as depicted in Figure 1 (C).

Figure 1. Distribution of message count per patient or clinician (A and D), total number of patient-sent or clinician-sent messages and count of unique patients or clinicians by year (B and E), and box-whisker graph of message count per patient or clinician by year (C and F).



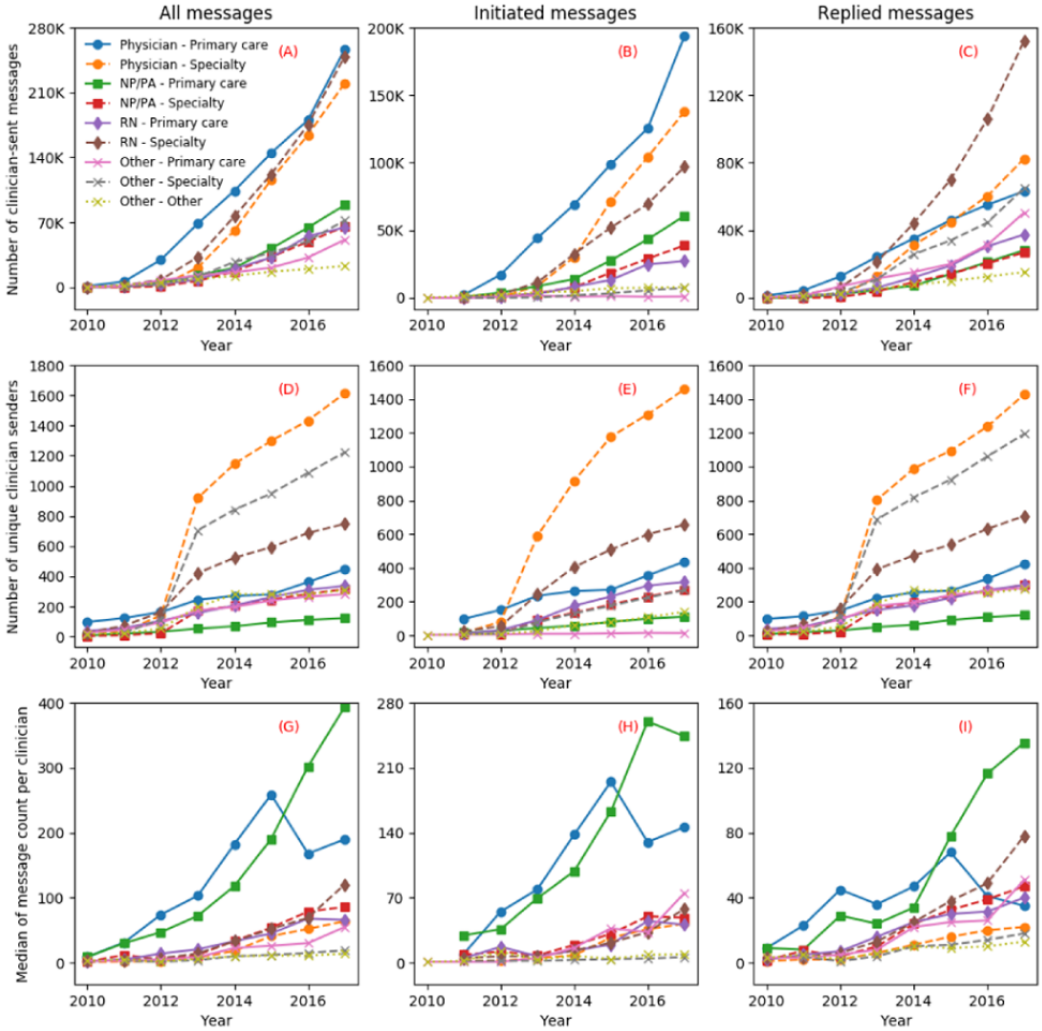
Clinician-Sent Messages

Figure 1 (D-F) depicts the distribution of message count per clinician, the total numbers of clinician-sent messages and unique clinicians by year, and the box-whisker graph of message count per clinician by year. In Figure 1 (D), the distribution of message count per clinician exhibits a "long tail" pattern, where the median is 77 but the max stretches far right at 18,314 (way beyond the X axis labeling range). Both the numbers of clinicians and clinician-sent messages were also increasing over time with the growths of the numbers of patients and patient-sent messages (Figure 1 [E]). The Pearson correlation coefficient between the total numbers of clinician-sent messages and patient-sent messages over years is r=1, suggesting that the total number of clinician-sent message is strongly associated with that of patient-sent message. The median and IQR of message count per clinician steadily increased during 2010-2017 (Figure 1 [F]). We observed that the count of clinicians largely increased but the median and IQR of message count per clinician decreased in 2013 compared with 2012 because many specialties at Mayo Clinic, Rochester started to use the patient portal.

Between 2010 and 2017, 19.93% (6773/8459) of the clinicians transitioned out of Mayo Clinic, Rochester and, therefore, we do not have information on their practice roles and care settings. We grouped the remaining 6773 clinicians, who generated 86.87% (3,071,529/3,535,741) messages, into 9 clinician categories. Based on these messages, we then analyzed the workload, message counts per clinician, and the distribution of messaging time in a day for each clinician category.

Figure 2 shows the total number of clinician-sent messages, the total number of clinicians, and the median of message count per clinician in each clinician category for each year. As shown in Figure 2 (A), the number of messages sent by clinicians in each clinician category was progressively increasing over time. Three clinician categories, namely "physician—primary care," "Registered Nurse (RN)—specialty," and "physician—specialty" had the largest increase in the generated messages. After 2014, the number of messages sent by these 3 clinician categories were over 2.5 times more than those of the other clinician categories combined. The category "physician—primary care" initiated the largest number of messages every year (Figure 2 [B]). During 2010-2013, the "physician—primary care" category also replied the most to the messages from patients, but after 2013, "RN—specialty" took over the top spot in responding to patients (Figure 2 [C]).

Figure 2. Total number of clinician-sent messages (A-C), the number of unique clinician senders (D-F), and median of message count per clinician (G-I) in each clinician category over years. NP: nurse practitioner; PA: physician assistant; RN: registered nurse.



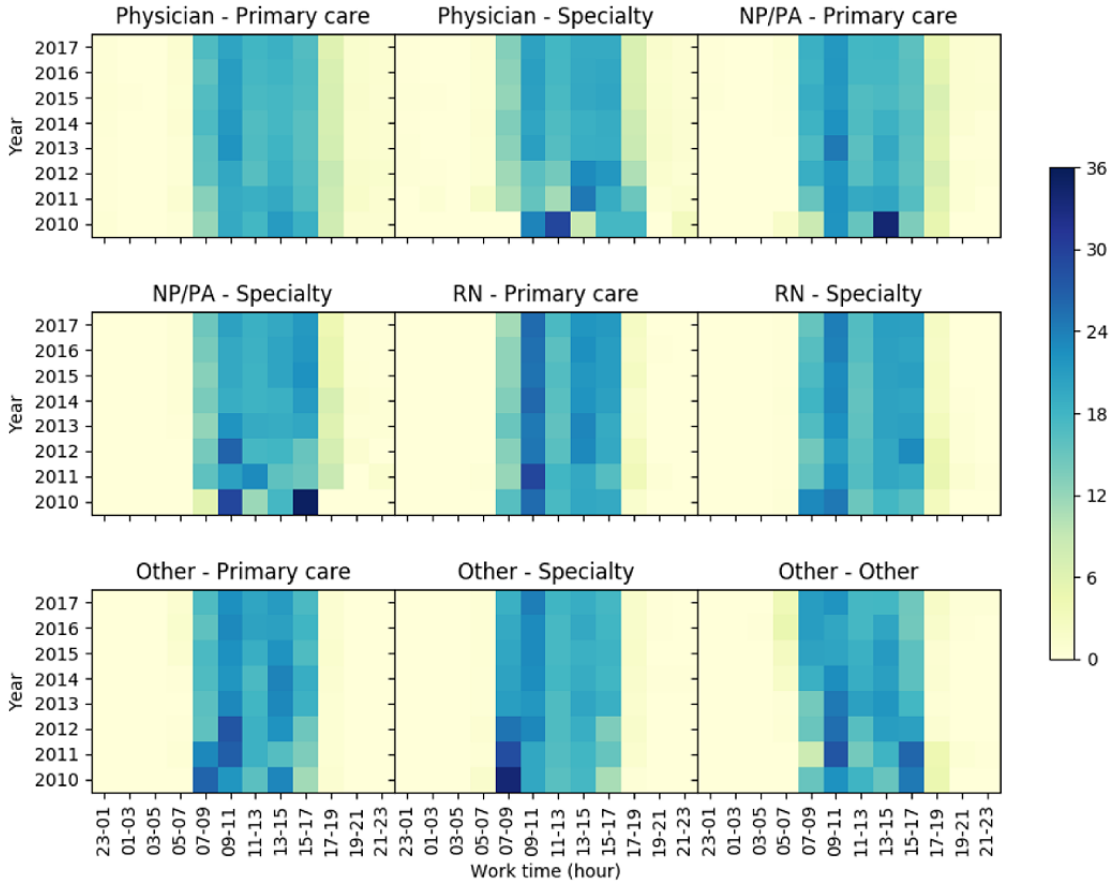
The total numbers of unique clinician senders in each category had a steady increase during 2010-2017 (Figure 2 [D-F]). After 2012, the clinician categories with top 3 message senders were "physician—specialty," "other—specialty," and "RN—specialty." Most of the clinicians in the "other—specialty" category were responsible for responding to patient messages rather than initiating messages. The "nurse practitioner (NP) or physician assistant (PA)—primary care" category has the smallest size of clinician senders (Figure 2 [D]).

In Figure 2 (G), the median of message count per clinician in 7 out of the 9 clinician categories steadily increased during 2010-2017. Among them, the top 2 clinician categories were "NP/PA—primary care" and "physician—primary care." In 2017, the median of message count per clinician by the "NP/PA—primary care" category was over 2 times of any other groups (Figure 2 [G-I]).

Clinician Messaging Workload Within a Day

We analyzed the clinician messaging time during a day across the 9 clinician categories in terms of message percentage (Figure 3) and message number (Figure S5-7, Multimedia Appendix 1). Most of the messages were sent from 7 AM to 5 PM. However, physicians and NP/PA also sent a considerable number of messages to patients between 5 PM and 7 PM. For example, the percentage of messages sent by "physician—primary care" during 5-7 PM increased from 7.95% (102/1283) in 2010 to 8.98% (2669/29,705) in 2012 and then dropped to 6.12% (15,714/256,761) in 2017. The percentage of messages sent by "NP/PA—primary care" during 5-7 PM remained constant around 5.08-6.75% (3/59-2,824/41,833), between 2010 and 2017.

Figure 3. Distribution of clinician messaging time in a day. Each color block represents the percentage of messages sent in the corresponding time slice during a day. NP: nurse practitioner; PA: physician assistant; RN: registered nurse.



Thread-Oriented Analysis

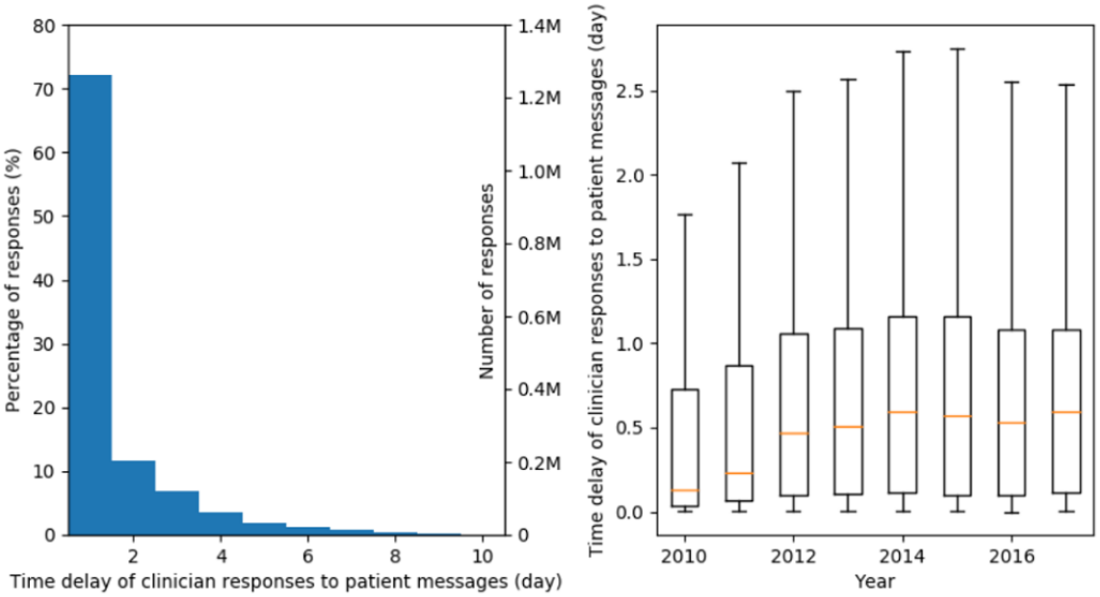
We identified 1,576,205 message threads in 3,887,542 messages, which include 1,332,931 patient-initiated messages and 243,274 clinician-initiated messages, for thread-oriented analysis. Among these message threads, we identified 1,761,739 clinician responses to patient messages.

Time Delay Between Patient Messages and Clinician Responses

We illustrated the distribution of time delay between clinician responses and patient messages in Figure 4. It appears that 96.0% (1,691,733/1,761,739) of patient messages were responded to by clinicians within 5 days. The shortest time a clinician spent responding to patients was 1 second, and the median was 0.6 days. With the increase in the total number of clinician messages and message count per clinician, the median time delay between clinician responses and patient messages

increased from 0.13 days in 2010 to 0.59 days in 2014 but remained steady (0.53-0.59 days) after 2014. The IQR of time delay showed a trend similar to the median time delay over years.

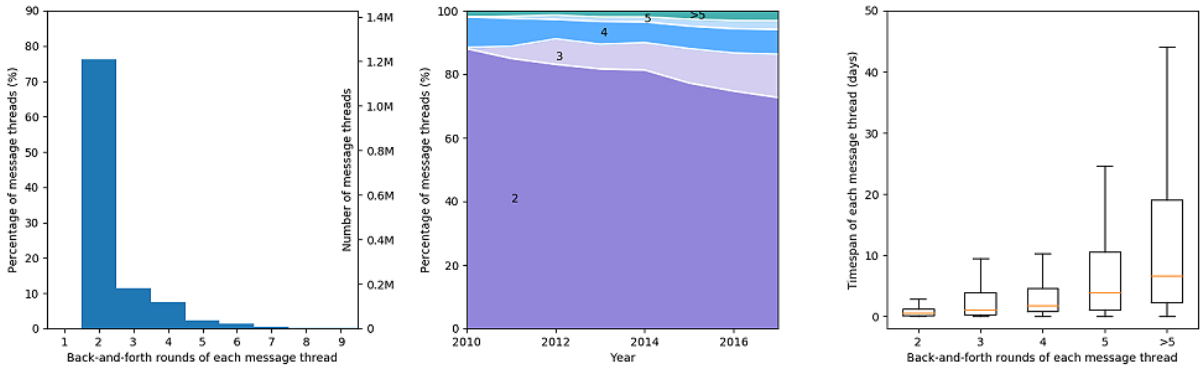
Figure 4. Distribution (left) and box-whisker graph (right) of time delay in the clinician responses to patient messages.



Back-and-Forth Messages Between Patients and Clinicians in a Message Thread

We analyzed the distribution of message thread lengths (Figure 5). We found that 95.1% (1,499,316/1,576,205) message threads had fewer than 5 back-and-forth messages. The median and maximum lengths of the message threads were 2 and 34, respectively. Between 2010 and 2017, the percentage of message threads with a length of 2 decreased from 88.0% (3429/3895) to 72.7% (372,424/512,395). The percentage of message threads with a length of 3 increased from 0.4% (16/3895) to 13.6% (69,754/512,395). The percentage of message threads with a length of 4 remained relatively stable, 6.1-9.6% (2354/38,537-374/3895). The percentage of the message threads with a length of 5 or more showed an uptrend over time. The median timespan of message threads was elongated as the message thread lengths increased.

Figure 5. Distribution (left), percentage (middle), and timespan (right) of message threads in terms of back-and-forth rounds in the message threads.



Discussion

Principal Findings

Secure messaging in patient portal provides patients more convenient access for more personalized medical advice and information [5,24]. Patients can send secure messages to their clinicians anywhere and anytime as they prefer, which fulfills a growing consumer model of medical care [14]. Similar to previous results on secure messaging [14-17], our patient portal experienced a large increase in the numbers of patients, clinicians, and secure messages during 2010-2017. We observed a strong correlation between patient count and patient message count. The growth of patients and messages per patient both contributes to the upsurge of patient messages, suggesting a strong adoption of this health care technique and a growing pattern of patient engagement. The significant volume difference between patient-initiated messages and patient-replied messages (1.5 million vs 0.5 million) indicates the behavioral variation of patients for initiating and replying to messages; in general,

the patients initiated a message when they encountered health issues instead of following an existing message thread. This implies potential opportunities to engage patients in communication and increase responses in pertinent contexts.

The clinician message volume is strongly associated with patient message volume because the clinicians were obligated to communicate with the patients for providing health care support. There were also practice-specific and role-specific patterns for the clinicians. For example, "physician—primary care" initiated the largest number of messages to patients every year, indicating their major role in patient engagement. After 2013, "RN—specialty" replied the most of messages to patients, suggesting that the patient portal opened up a longed-for channel of taking patient information needs in specialty care [25], and that specialty RNs represented the first-line taskforce to accommodate the needs.

The patient portal also manifested upward trends in the ratio of patient count to clinician count and numbers of messages per patient and clinician. The findings suggest an inflation of clinician workload in the secure messaging. Specifically, the "physician—primary care" and "NP/PA—primary care" categories had the heaviest per-clinician messaging workload every year because the "physician—primary care" category generated the largest number of clinician messages and the "NP/PA—primary care" category has the smallest size of clinician senders. This indicates where optimization is most needed in easing the clinician workload. Most (93.2%, 2,862,978/3,071,529) of the clinician messages were sent from 7 AM to 5 PM but 17.3% (532,894/3,071,529) of the messages were sent by clinicians to patients around noon (11 AM to 1 PM), and 5.94% (182,468/3,071,529) were sent after 5 PM (5-11 PM). Our findings were consistent with those from previous studies that secure messaging has a negative effect on clinician workload [12], and that about 24% of the work is carried out by clinicians on electronic health records for non–face-to-face care after clinician hours [13].

Most of the clinicians responded to the patients in a timely manner. The median time delay of clinician responses to patient message remained constant (less than 0.60 days). Although the message volume and clinician workload persistently increased over years, the clinicians managed to respond to the patients in time. Regarding the message threads, the patients usually communicated with the clinicians in a few back-and-forth rounds. Possibly, patients and clinicians communicated via secure messaging for noncomplicated scenarios, or clinicians made a timely decision on a certain back-and-forth round about what acute issues and complex situations will require face-to-face visit or follow-up. The percentage of message threads with more than 2 rounds was rising, which may imply that the patients and clinicians were increasingly comfortable to communicate about more complex situations. However, further investigation is necessary to understand the potential mechanism of the messaging complexity of a communication thread.

The growing portal messages had different indications or impacts on its stakeholders. For the patients, a majority of them in this study received timely response to their concerns and requests. In the future, the patient engagement would probably continue to increase in terms of patient count and message count per patient, in particular, during and after the COVID-19 pandemic. For the clinicians, most of them think that secure messaging can have a positive effect on quality of care and patient safety, but secure messaging increased their burden of indirect patient care, as noted by Hoonakker et al [12]. In some cases, clinicians need to send messages to patients after 5 PM in order to respond to them in time. With the rise of patient engagement and message volume over time [14-17], it will be critical to properly distribute the communication load for better efficiency and for avoiding clinician burnouts. For the health care system, patient portals and secure messaging may have a favorable impact on the cost-effectiveness of care [26]. They could help to not only cut the administrative cost by alleviating the operational burden [27], but also reduce patients' health care utilization by improving their functional status [28]. However, the detailed economical or cost-effectiveness analysis is beyond the scope of this paper, due to the difficulty of collecting relevant data. In the near future, some other health care systems will probably confront the management challenge of the fast-growing volume of patient messages [12]. These health care systems would require new policies, billing models, or additional infrastructures. For example, more NP/PA, RN, and other support staff would be involved in replying to the increasing volume of patient messages. The artificial intelligence (AI) and natural language processing (NLP) tools would even be invested and developed to support the care teams for secure messaging [29-31].

Our study has several limitations. First, the secure messages were collected from Mayo Clinic, Rochester, a tertiary care institution for complex medical conditions, and might not be representative of different patient populations or clinical settings in other parts of the country. Second, the patient secure messages after 2017 were not included in this study due to the upgrade of the patient portal system at Mayo Clinic in 2018. The COVID-19 pandemic is transforming the health care delivery via telehealth, and we would investigate the secure messaging after 2017 and the impact of the COVID-19 on the patient portal system for a future study. Third, the 9 clinician categories we used were not always accurate because a small portion of clinicians changed their practice roles and care settings during the study period. Fourth, there was no guarantee that the clinician-initiated or clinician-replied messages were always written by themselves. Finally, linking patient secure messages to other patient medical records for a more detailed study of the needs of different patient populations is out of the scope of this study, but represents an important area we would like to explore in the future.

Conclusions

We performed message-oriented, sender-oriented, and thread-oriented analyses to probe and characterize millions of secure medical messages generated by patients and clinicians with diverse backgrounds. We analyzed the message volumes, patient or clinician population sizes, message counts per patient or clinician, and their trends over years. We computed the time distribution of clinician messaging to further understand their workload in different time slices of a day. For each message

thread, we calculated the time delay between patient messages and clinician responses to examine the responding efficiency and the number of back-and-forth rounds to roughly assess the communication complexity.

Our study shows a steady rise in patient involvement, through the use of secure messaging, and workload per clinician over years. However, most clinicians were responding to the patients in a timely manner in order to meet their needs. Our findings shed light on opportunities for care teams to improve messaging tasks and optimize clinician workload and for the experts in AI and NLP to develop robust and intelligent messaging tools to support the care teams for better communication efficiency and quality. These findings offer valuable information on the digital interaction between patients and clinicians and may serve as a reference for promoting patient-centered care.

In the future, we will perform a content analysis of patient secure messages using AI and NLP and examine the patient populations in terms of socioeconomic factors by linking them to patient medical records. We will also perform a comparative study between patient portal messages and traditional health care services and a survey study to understand patient and clinician experiences on using patient portal messaging.

Acknowledgments

Funding for this study was provided by the Mayo Clinic Center for Clinical and Translational Science (UL1TR002377) and the National Library of Medicine (5K01LM012102).

Authors' Contributions

LY designed and guided the research study, discussed the methods and results, and revised the manuscript. MH preprocessed the data, implemented the algorithms, performed the computations and analyses, and drafted and revised the manuscript. JF discussed the methods and results and revised the manuscript. JP, NDS, and BAC participated in the discussion and provided feedback on the manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary materials.
[DOCX File , 930 KB - jmir_v24i1e17273_app1.docx]

References

1. Irizarry T, DeVito Dabbs A, Curran CR. Patient Portals and Patient Engagement: A State of the Science Review. J Med Internet Res 2015 Jun 23;17(6):e148 [FREE Full text] [doi: 10.2196/jmir.4255] [Medline: 26104044]
2. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongsa N, Pasick R, et al. Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. J Med Internet Res 2015 Dec 03;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]
3. Centers for Medicare & Medicaid Services (CMS), HHS. Medicare and Medicaid programs; electronic health record incentive program--stage 2. Final rule. Fed Regist 2012 Sep 04;77(171):53967-54162 [FREE Full text] [Medline: 22946138]
4. Henry J, Barker W, Kachay L. Electronic Capabilities for Patient Engagement among U.S. Non-Federal Acute Care Hospitals: 2012-2015. The Office of the National Coordinator for Health Information Technology. 2019. URL: https://www.healthit.gov/sites/default/files/briefs/2015_patient_engagement_data_brief.pdf [accessed 2021-12-10]
5. Kruse CS, Bolton K, Freriks G. The effect of patient portals on quality outcomes and its implications to meaningful use: a systematic review. J Med Internet Res 2015 Feb 10;17(2):e44 [FREE Full text] [doi: 10.2196/jmir.3171] [Medline: 25669240]
6. Wade-Vuturo AE, Mayberry LS, Osborn CY. Secure messaging and diabetes management: experiences and perspectives of patient portal users. J Am Med Inform Assoc 2013 May 01;20(3):519-525 [FREE Full text] [doi: 10.1136/amiajnl-2012-001253] [Medline: 23242764]
7. Osborn CY, Mayberry LS, Wallston KA, Johnson KB, Elasy TA. Understanding patient portal use: implications for medication management. J Med Internet Res 2013 Jul 03;15(7):e133 [FREE Full text] [doi: 10.2196/jmir.2589] [Medline: 23823974]
8. Horvath M, Levy J, L'Engle P, Carlson B, Ahmad A, Ferranti J. Impact of health portal enrollment with email reminders on adherence to clinic appointments: a pilot study. J Med Internet Res 2011 May 26;13(2):e41 [FREE Full text] [doi: 10.2196/jmir.1702] [Medline: 21616784]
9. Guy R, Hocking J, Wand H, Stott S, Ali H, Kaldor J. How effective are short message service reminders at increasing clinic attendance? A meta-analysis and systematic review. Health Serv Res 2012 Apr 08;47(2):614-632 [FREE Full text] [doi: 10.1111/j.1475-6773.2011.01342.x] [Medline: 22091980]

10. Zhou YY, Garrido T, Chin HL, Wiesenthal AM, Liang LL. Patient access to an electronic health record with secure messaging: impact on primary care utilization. Am J Manag Care 2007 Jul;13(7):418-424 [FREE Full text] [Medline: 17620037]
11. Simon GE, Ralston JD, Savarino J, Pabiniak C, Wentzel C, Operskalski BH. Randomized trial of depression follow-up care by online messaging. J Gen Intern Med 2011 Jul 8;26(7):698-704 [FREE Full text] [doi: 10.1007/s11606-011-1679-8] [Medline: 21384219]
12. Hoonakker PL, Carayon P, Cartmill RS. The impact of secure messaging on workflow in primary care: Results of a multiple-case, multiple-method study. Int J Med Inform 2017 Apr;100:63-76 [FREE Full text] [doi: 10.1016/j.ijmedinf.2017.01.004] [Medline: 28241939]
13. Arndt BG, Beasley JW, Watkinson MD, Temte JL, Tuan W, Sinsky CA, et al. Tethered to the EHR: Primary Care Physician Workload Assessment Using EHR Event Log Data and Time-Motion Observations. Ann Fam Med 2017 Sep 11;15(5):419-426 [FREE Full text] [doi: 10.1370/afm.2121] [Medline: 28893811]
14. Crotty BH, Tamrat Y, Mostaghimi A, Safran C, Landon BE. Patient-to-physician messaging: volume nearly tripled as more patients joined system, but per capita rate plateaued. Health Aff (Millwood) 2014 Oct;33(10):1817-1822 [FREE Full text] [doi: 10.1377/hlthaff.2013.1145] [Medline: 25288428]
15. Cronin R, Davis S, Shenson J, Chen Q, Rosenbloom S, Jackson G. Growth of Secure Messaging Through a Patient Portal as a Form of Outpatient Interaction across Clinical Specialties. Appl Clin Inform 2017 Dec 19;06(02):288-304. [doi: 10.4338/aci-2014-12-ra-0117]
16. Shenson JA, Cronin RM, Davis SE, Chen Q, Jackson GP. Rapid growth in surgeons' use of secure messaging in a patient portal. Surg Endosc 2016 Apr 27;30(4):1432-1440 [FREE Full text] [doi: 10.1007/s00464-015-4347-y] [Medline: 26123340]
17. Masterman M, Cronin R, Davis S, Shenson J, Jackson G. Adoption of Secure Messaging in a Patient Portal across Pediatric Specialties. AMIA Annu Symp Proc 2017;2016:1930-1939 [FREE Full text] [Medline: 28269952]
18. Wosik J, Fudim M, Cameron B, Gellad Z, Cho A, Phinney D, et al. Telehealth transformation: COVID-19 and the rise of virtual care. J Am Med Inform Assoc 2020 Jun 01;27(6):957-962 [FREE Full text] [doi: 10.1093/jamia/ocaa067] [Medline: 32311034]
19. Patient Online Services. 2021. URL: https://onlineservices.mayoclinic.org/content/staticpatient/showpage/patientonline [accessed 2021-12-03]
20. McKinney W. Data structures for statistical computing in python. 2010 Jun 1 Presented at: Proceedings of the 9th Python in Science Conference; 2010; Austin, Texas. [doi: 10.25080/Majora-92bf1922-00a]
21. Harris CR, Millman KJ, van der Walt SJ, Gommers R, Virtanen P, Cournapeau D, et al. Array programming with NumPy. Nature 2020 Sep 16;585(7825):357-362 [FREE Full text] [doi: 10.1038/s41586-020-2649-2] [Medline: 32939066]
22. Virtanen P, Gommers R, Oliphant T, Haberland M, Reddy T, Cournapeau D, SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. Nat Methods 2020 Mar;17(3):261-272 [FREE Full text] [doi: 10.1038/s41592-019-0686-2] [Medline: 32015543]
23. Hunter JD. Matplotlib: A 2D Graphics Environment. Comput. Sci. Eng 2007 May;9(3):90-95. [doi: 10.1109/mcse.2007.55]
24. Coughlin S, Prochaska J, Williams LB, Besenyi G, Heboyan V, Goggans S, et al. Patient web portals, disease management, and primary prevention. RMHP 2017 Apr;10:33-40. [doi: 10.2147/rmhp.s130431]
25. Chen AH, Yee HF. Improving the primary care-specialty care interface: getting from here to there. Arch Intern Med 2009 Jun 08;169(11):1024-1026. [doi: 10.1001/archinternmed.2009.140] [Medline: 19506170]
26. Riippa I, Linna M, Rönkkö I. A Patient Portal With Electronic Messaging: Controlled Before-and-After Study. J Med Internet Res 2015 Nov 09;17(11):e250 [FREE Full text] [doi: 10.2196/jmir.4487] [Medline: 26553595]
27. Kamo N, Bender AJ, Kalmady K, Blackmore CC. Meaningful use of the electronic patient portal - Virginia Mason's journey to create the perfect online patient experience. Healthc (Amst) 2017 Dec;5(4):221-226. [doi: 10.1016/j.hjdsi.2016.09.003] [Medline: 27727028]
28. Sorondo B, Allen A, Fathima S, Bayleran J, Sabbagh I. Patient Portal as a Tool for Enhancing Patient Experience and Improving Quality of Care in Primary Care Practices. EGEMS (Wash DC) 2016 Jan 26;4(1):1262 [FREE Full text] [doi: 10.13063/2327-9214.1262] [Medline: 28203611]
29. Sulieman L, Gilmore D, French C, Cronin RM, Jackson GP, Russell M, et al. Classifying patient portal messages using Convolutional Neural Networks. J Biomed Inform 2017 Oct;74:59-70 [FREE Full text] [doi: 10.1016/j.jbi.2017.08.014] [Medline: 28864104]
30. Sulieman L, Robinson JR, Jackson GP. Automating the Classification of Complexity of Medical Decision-Making in Patient-Provider Messaging in a Patient Portal. J Surg Res 2020 Nov;255:224-232 [FREE Full text] [doi: 10.1016/j.jss.2020.05.039] [Medline: 32570124]
31. De A, Huang M, Feng T, Yue X, Yao L. Analyzing Patient Secure Messages Using a Fast Health Care Interoperability Resources (FIHR)-Based Data Model: Development and Topic Modeling Study. J Med Internet Res 2021 Jul 30;23(7):e26770 [FREE Full text] [doi: 10.2196/26770] [Medline: 34328444]

Abbreviations

AI: artificial intelligence
ID: identifier
NLP: natural language processing
NP: nurse practitioner
PA: physician assistant
RN: registered nurse

Edited by R Kukafka; submitted 13.05.21; peer-reviewed by A Barker, J Ritchie, G Cenikj; comments to author 28.06.21; revised version received 23.08.21; accepted 18.11.21; published 11.01.22.

Please cite as:
Huang M, Fan J, Prigge J, Shah ND, Costello BA, Yao L
Characterizing Patient-Clinician Communication in Secure Medical Messages: Retrospective Study
J Med Internet Res 2022;24(1):e17273
URL: https://www.jmir.org/2022/1/e17273
doi:10.2196/17273
PMID:35014964

Original Paper

Predictors of Smartphone and Tablet Use Among Patients With Hypertension: Secondary Analysis of Health Information National Trends Survey Data

Chinwe E Eze[1], BPharm, MS; Brady T West[2], PhD; Michael P Dorsch[1], PharmD, MS; Antoinette B Coe[1], PharmD, PhD; Corey A Lester[1], PharmD, PhD; Lorraine R Buis[3], PhD; Karen Farris[1], PhD

[1]College of Pharmacy, University of Michigan, Ann Arbor, MI, United States

[2]Institute for Social Research, University of Michigan, Ann Arbor, MI, United States

[3]Department of Family Medicine, University of Michigan, Ann Arbor, MI, United States

Corresponding Author:
Chinwe E Eze, BPharm, MS
College of Pharmacy
University of Michigan
428 Church Street
Ann Arbor, MI, 48109-1065
United States
Phone: 1 7346806587
Email: ceeze@umich.edu

Abstract

Background: Uncontrolled hypertension leads to significant morbidity and mortality. The use of mobile health technology, such as smartphones, for remote blood pressure (BP) monitoring has improved BP control. An increase in BP control is more significant when patients can remotely communicate with their health care providers through technologies and receive feedback. Little is known about the predictors of remote BP monitoring among hypertensive populations.

Objective: The objective of this study is to quantify the predictors of smartphone and tablet use in achieving health goals and communicating with health care providers via SMS text messaging among hypertensive patients in the United States.

Methods: This study was a cross-sectional, secondary analysis of the 2017 and 2018 Health Information National Trends Survey 5, cycles 1 and 2 data. A total of 3045 respondents answered "Yes" to the question "Has a doctor or other healthcare provider ever told you that you had high blood pressure or hypertension?", which defined the subpopulation used in this study. We applied the Health Information National Trends Survey full sample weight to calculate the population estimates and 50 replicate weights to calculate the SEs of the estimates. We used design-adjusted descriptive statistics to describe the characteristics of respondents who are hypertensive based on relevant survey items. Design-adjusted multivariable logistic regression models were fitted to estimate predictors of *achieving health goals with the help of smartphone or tablet* and *sending or receiving an SMS text message to or from a health care provider in the last 12 months*.

Results: An estimated 36.9%, SE 0.9% (183,285,150/497,278,883) of the weighted adult population in the United States had hypertension. The mean age of the hypertensive population was 58.3 (SE 0.48) years. Electronic communication with the doctor or doctor's office through email or internet (odds ratio 2.93, 95% CI 1.85-4.63; *P*<.001) and having a wellness app (odds ratio 1.82, 95% CI 1.16-2.86; *P*=.02) were significant predictors of using SMS text message communication with a health care professional, adjusting for other demographic and technology-related variables. The odds of achieving health-related goals with the help of a tablet or smartphone declined significantly with older age (*P*<.001) and ownership of basic cellphones (*P*=.04). However, they increased significantly with being a woman (*P*=.045) or with being married (*P*=.03), having a wellness app (*P*<.001), using devices other than smartphones or tablets to monitor health (*P*=.008), making health treatment decisions (*P*=.048), and discussing with a provider (*P*=.02) with the help of a tablet or smartphone.

Conclusions: Intervention measures accounting for age, gender, marital status, and the patient's technology-related health behaviors are required to increase smartphone and tablet use in self-care and SMS text message communication with health care providers.

(J Med Internet Res 2022;24(1):e33188) doi:10.2196/33188

KEYWORDS

hypertension; mHealth; remote monitoring; telemonitoring; smartphones; tablets; text messaging; Health Information National Trends Survey; mobile health; digital health; mobile phone

Introduction

Background

Among the 121.5 million adults in the United States with hypertension, 61.2% are aware of their disease condition, and 50.4% are receiving treatment, but only about 22% have their blood pressure (BP) controlled [1]. Uncontrolled hypertension can lead to stroke [2], systemic embolism and bleeding [3], congestive heart failure [4], myocardial infarction [5], renal damage, dementia, aortic aneurysm, angina pectoris, metabolic syndrome, diabetes, blindness, and death [6,7]. The 2021 Heart Disease and Stroke statistics report that 57.2% of all deaths recorded in the United States from 2008 to 2018 were attributed to hypertension [1]. Despite effective lifestyle and pharmaceutical treatments, the number of patients with uncontrolled BP in the United States is undesirable. Thus, there is a need to harness every possible arsenal to mitigate this challenge.

One strategy to improve BP control involves patients in their disease management through technology [8]. Recent innovations in information and communication technology provide excellent opportunities for improvements in hypertension control. There has been a steady increase in internet users and mobile cellular subscribers since 2000 [9]. According to a 2021 Pew Research Center report, 93% of adult Americans now use the internet, and an increase in internet use is seen across all age groups [9]. Moreover, 97% of adult Americans own a cellphone, and 85% now use smartphones [10].

In considering technology and BP control, patients with hypertension can now measure their BP using electronic monitors and transmit the results to their health provider through electronic health record platforms on their smartphones, tablets, or computers, and get feedback through the same channels without having to leave the comfort of their homes [11]. Phone calls, SMS text message alerts, health apps, emails, and alarms have also been used, and collectively this is called telemonitoring. Improvements in BP control have been noted with this type of remote monitoring. For example, a pharmacist-led telemonitoring intervention involving weekly electronic transmission of home-measured BP and regulated telephone visits among 450 patients with uncontrolled BP resulted in a significant decrease in systolic BP at 6, 12, and 18 months of −10.7 mm Hg (95% CI −14.3 to −7.3 mm Hg), P<.001; −9.7 mm Hg (95% CI −13.4 to −6.0 mm Hg), P<.001; and −6.6 mm Hg (−10.7 to −2.5 mm Hg), P=.004, respectively [11]. In addition, this study reported an increase in the proportion of patients with controlled BP in the telemonitoring group (71.8%, 95% CI 65.0-77.8) compared with the usual care group (57.1%, 95% CI 51.5-62.6) [12]. More generally, the use of SMS text messages as reminders and health education delivery led to improvements in behavior changes, hypertension knowledge, medication adherence, and BP among patients with hypertension [13-16]. A meta-analysis of 46 randomized controlled trials reported that home BP telemonitoring decreased systolic BP −3.99 mm Hg (95% CI −5.06 to −2.93; P<.001) and diastolic BP −1.99 mm Hg (95% CI −2.60 to −1.39; P<.001) in the intervention groups compared with usual care [17]. However, these are mostly intervention studies that are not nationally representative.

Although we know the advantages of these technologies in achieving favorable health outcomes, little is known about the predictors of their use among patients with hypertension. Using the Health Information National Trends Survey (HINTS), Langford et al [18] examined the prevalence of smartphones, basic phones, and tablets and compared respondents who are hypertensive and nonhypertensive. They found that 68%, 55%, and 16% of the hypertensive population had smartphones, tablets, and basic mobile phones, respectively. Younger respondents who are hypertensive were more likely to own a smartphone or tablet and have a health-related app. The ownership of smartphones or tablets increased with an increase in educational attainment. Another HINTS study focused on respondents with one or more chronic medical conditions and found that gender, age, employment status, and having a health app were associated with achieving a health-related goal with a smartphone or tablet. However, this study did not differentiate the respondents according to the disease conditions in the analysis [19]. Other studies on mobile health app use did not focus on people with hypertension [20,21]. Therefore, there is a need for more hypertension-focused studies to identify the factors that impact mobile health (mHealth) technology use among this patient population.

Objectives

The aim of this study is to quantify the predictors of smartphone and tablet use in achieving health goals and communicating with health care providers via SMS text messaging among patients with hypertension. Our research question was, “What are the relationships of patients’ characteristics with the use of a smartphone or tablet to achieve health goals and sending or receiving text messages to or from healthcare professionals, among a nationally representative sample with hypertension?” This study provides nationally representative estimates regarding the predictors of using a smartphone or tablet to achieve health-related goals and SMS text messaging communication with health care professionals among respondents who are hypertensive. It also illuminates respondents’ factors associated with the use of these communication approaches. This will help us identify where and how to channel efforts to improve involvement of patients in telemonitoring of BP when health care providers work with their patients to increase smartphone and tablet use for health services. These results will also inform our questions for further studies to understand patients’ experiences with technology for BP control.

Methods

Design

This study was a cross-sectional, secondary quantitative analysis of the 2017 and 2018 HINTS 5, cycles 1 and 2 data. We combined the 2 cycles to provide more robust estimates of our relationships of interest. The study was considered exempt by the University of Michigan institutional review board (approval number: HUM00208364).

Data Collection

The HINTS was developed by the Health Communication and Informatics Research branch of the National Cancer Institute. It is a publicly available, nationally representative survey that monitors how American adults aged ≥18 years obtain and use health information. HINTS has been carried out every few years since 2003, and the target population is adult Americans aged ≥18 years in the civilian noninstitutionalized population of the United States. HINTS uses a 2-stage sampling design, and residents in high minority strata are oversampled. A high minority stratum represents places with ≥34% Hispanic or African Americans. The data had both a full sample weight and 50 replicate weights assigned to each completed questionnaire for the adult sample. The 50 replicate weights were computed using the jackknife replication method. The full sample weight enables the calculation of population and subpopulation estimates, whereas the 50 replicate weights allow for the analysis of design-adjusted SEs for these estimates. The sample weights allow valid inferences from the responding sample to the population, accounting for unequal probability of selection, nonresponse, and noncoverage biases. The details of the sampling methods and weighting approaches are available in the HINTS 5, cycles 1 and 2 methodology reports [22,23].

Participants or Sample Size

A total of 6789 respondents completed the HINTS 5 cycles 1 and 2 questionnaires. Respondents to HINTS who answered "*Yes*" to the question "Has a doctor or other health provider ever told you that you had high blood pressure or hypertension?" were the subpopulations used in this study. Out of the 6789 respondents, 3045 (44.85%) belonged to this subpopulation and thus constituted the final sample included in this analysis.

Variables of Interest

The dependent variables were (1) Has your tablet or smartphone helped you track progress on a health-related goal, such as quitting smoking, losing weight, or increasing physical activity? (yes or no) and (2) Have you sent or received an SMS text message from a doctor or other health care professional within the last 12 months? (yes, no, or don't know). The *no* and *don't know* responses were combined to a single *no* response for logistic regression analysis. In this study, we described and predicted these variables. We selected these 2 items because they are most closely related to the concept of telemonitoring of BP. We also provided population proportion estimates of the following variables: Has your tablet or smartphone helped you make a decision about how to treat an illness or condition? (yes or no); Has your tablet or smartphone helped you in discussions with your health care provider? (yes or no); Other than a tablet or smartphone, have you used an electronic device to monitor or track your health within the last 12 months? (yes or no); Have you shared health information from either an electronic monitoring device or smartphone with a health professional within the last 12 months? (yes or no); and in the past 12 months, have you used a computer, smartphone, or other electronic means to use email or the internet to communicate with a doctor or doctor's office? (yes or no).

The independent variables included respondents' demographics (such as age, educational level, marital status, and income) and clinical characteristics (BMI, comorbidities, and general health status). Technology-related covariates included technology access, such as ownership of smartphones, tablets, wellness health apps, and basic cellphones. Technology-related behaviors, such as electronic communication with the doctor or doctor's office through email or the internet were also included. These covariates were selected as they are technology-related items that can be applied to BP telemonitoring.

Statistical Analysis

We accounted for the sampling weights and complex sample design features in all analyses to obtain population-level estimates for the United States using the R *survey* package by Thomas Lumley. Variance estimates were computed using the jackknife replication method, and specialized (unconditional) subpopulation analyses were not required when using this replication approach [24]. Descriptive statistics were used to analyze the characteristics of the respondents based on relevant demographics and covariates. We fit multivariable logistic regression models to the variables of interest to determine the most important predictors of the dependent variables. We first used demographic variables only and then tested the full model with clinical and technology use variables. The pseudo maximum likelihood estimation method was used to fit the regression models. To arrive at the final fitted model, we used a step-by-step approach starting from the preliminary bivariate analyses of potential predictors, followed by fitting different models containing all the anticipated predictors and variables of interest as well as interaction terms. We used the *regTermTest* function in the *survey* package to test the significance of the predictors using the design-adjusted Wald tests. None of the interaction terms was found to be significant. We identified the best-fitting model by choosing the model with the lowest design-adjusted Akaike information criterion [25]. Some nonsignificant predictors were retained in the models because they were found to be associated with hypertension in previous studies [26-28] and removing them did not result in a better-fitting model. Statistical significance was set at $P \leq .05$. All analyses were conducted using the JJ Allaire R Studio (version 3.6.1).

Results

Demographics and Clinical Characteristics

Out of the 497,278,883 estimated weighted population surveyed, 183,285,150 (36.9%, SE 0.9%) responded "Yes" to having hypertension. The 183,285,150 estimated hypertensive population constituted the denominator for all analyses in this study. The mean age of the hypertensive population was 58.3

(SE 0.48) years. Among people with hypertension, there were more men (52.7%) than women (47.3%), and most persons were aged between 50 and 64 years (Table 1). The hypertensive population was predominantly non-Hispanic White people (66.9%), and most had some college education or more (61.2%). Most were married or living as married (57.1%), and more than three-quarters considered themselves to be in good, very good, or excellent health. Less than half of this subpopulation was employed (46.7%), and more than two-thirds earned a yearly household income below US $75,000. Diabetes was the most commonly reported comorbidity (33.7%).

Table 1. Design-adjusted estimates of demographics and clinical characteristics among the hypertensive population (sample size=3045; estimated population size=183,285,150).

Variable and category	Value	95% CI
Age (years), mean (SE)	58.3 (0.48)	57.31-59.21
Age groups (years), % (SE)		
18-34	6 (1)	4.1-8.0
35-49	22.2 (1.5)	19.2-25.2
50-64	37.6 (1.4)	34.8-40.3
65-74	18.8 (0.6)	17.6-19.9
≥75	15.4 (0.6)	14.2-16.6
Gender, % (SE)		
Male	52.7 (1.3)	50.1-55.3
Female	47.3 (1.3)	44.7-49.9
Education level, % (SE)		
Less than high school	10.8 (1)	8.8-12.8
High school graduate	28 (1.4)	25.3-30.8
Some college	37.4 (1.4)	34.6-40.2
College graduate or more	23.8 (1.0)	21.9-25.6
Race or ethnicity, % (SE)		
Non-Hispanic White	66.9 (1.1)	64.7-69.1
Non-Hispanic Black or African American	13.9 (0.8)	12.4-15.4
Hispanic	12.8 (0.9)	11.1-14.5
Non-Hispanic Asian	3.2 (0.5)	2.2-4.2
Non-Hispanic other	3.1 (0.4)	2.4-3.9
Marital status, % (SE)		
Married	54.7 (1.2)	52.3-57.1
Living as married	2.4 (0.4)	1.6-3.2
Divorced	11.4 (0.6)	10.2-12.3
Widowed	9.2 (0.6)	8.0-10.4
Separated	1.6 (0.3)	1.0-2.7
Never married	20.7 (1.3)	18.1-23.4
Household yearly income (US $), % (SE)		
<20,000	21.1 (1.2)	18.7-23.5
20,000 to 35,000	13.2 (0.8)	11.6-14.8
35,000 to <50,000	15.2 (1.1)	13.0-17.3
50,000 to <75,000	19.4 (1.2)	17.0-21.9
≥75,000	31.1 (1.2)	28.6-33.5
Employment status, % (SE)		
Employed	46.7 (1.6)	46.7-49.7
Unemployed	53.3 (1.6)	50.1-56.4
Smoked at least 100 cigarettes, % (SE)		
Yes	44.7 (1.6)	41.6-47.7
No	55.3 (1.6)	52.3-58.4
BMI, mean (SE)	31.1 (19.1)	30.7-31.4

Variable and category	Value	95% CI
General health, % (SE)		
Excellent	5.5 (0.6)	4.4-6.6
Very good	28 (1.5)	25.0-30.9
Good	42 (1.5)	38.9-45.0
Fair	20.3 (1.3)	17.9-22.8
Poor	4.3 (0.6)	3.1-5.4
Diabetes, % (SE)		
Yes	33.7 (1.4)	31.0-36.4
No	66.3 (1.4)	63.6-69.0
Heart condition, % (SE)		
Yes	15.8 (1)	13.8-17.7
No	84.2 (1)	82.3-86.2
Depression, % (SE)		
Yes	27.8 (1.3)	25.4-30.3
No	72.2 (1.3)	69.7-74.6

Ownership and Use of Electronic Devices

In the hypertensive subpopulation, the distribution of ownership of electronic devices was as follows: smartphones (69.4%); tablets (54.7%); and basic cellphones (21.8%; Table 2). Almost three-quarters (74%) had accessed the internet; however, lower proportions used their smartphones or tablets to achieve health-related goals (36.1%) and sent or received SMS text messages to or from their health care professionals (30%). Only one-third (33.6%) of the hypertensive population communicated electronically with their doctor or doctor's office through email or the internet.

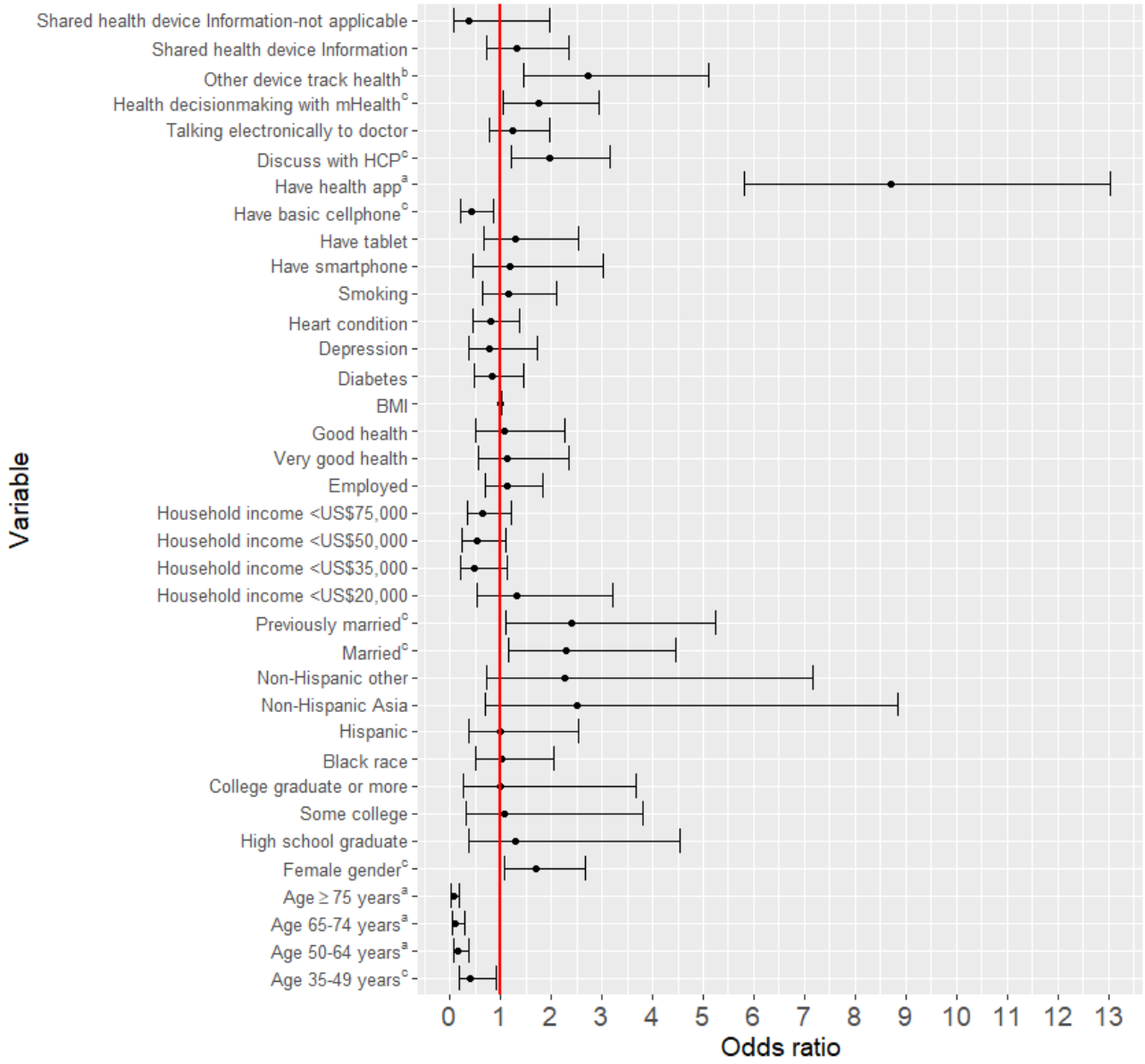
Table 2. Design-adjusted proportions for ownership and use of mobile health (mHealth) electronic devices among the hypertensive population (sample size=3045; estimated population size=183,285,150).

Variable and category	Value, n (%)	SE (%)	95% CI (%)
Technology access and use			
Have only basic cellphone			
Yes	39,974,491 (21.81)	1.10	19.65-23.96
No	143,310,659 (78.19)	1.10	76.04-80.35
Have smartphone			
Yes	127,108,252 (69.35)	1.36	66.69-72.01
No	56,176,898 (30.65)	1.36	28.00-33.31
Have tablet			
Yes	100,183,663 (54.66)	1.44	51.84-57.48
No	83,101,487 (45.34)	1.44	42.52-48.16
Use internet			
Yes	135,631,011 (74)	1.30	71.44-76.56
No	47,654,139 (26)	1.30	23.44-28.56
Have health apps			
Yes	74,395,442 (40.59)	1.76	37.15-44.04
No	108,889,708 (59.41)	1.76	55.96-62.85
Technology-related health behaviors			
Make health treatment decision with mHealth			
Yes	62,738,507 (34.23)	1.97	30.36-38.09
No	120,546,643 (65.77)	1.97	61.91-69.64
Discuss with health provider with help of tablet or smartphone			
Yes	61,730,439 (33.68)	1.51	30.72-36.64
No	121,554,711 (66.32)	1.51	63.36-69.28
Used other devices apart from tablet and smartphone to monitor or track health			
Yes	76,154,980 (41.55)	1.25	39.10-44.01
No	107,130,170 (58.45)	1.25	55.99-60.90
Shared health information from electronic device, tablet, or smartphone with health provider			
Yes	38,838,123 (21.19)	1.02	19.19-23.19
No	129,875,857 (70.86)	1.42	68.07-73.64
Not applicable	14,571,170 (7.95)	0.88	6.23-9.68
Electronic communication with doctor or doctor's office via email or internet			
Yes	61,620,467 (33.62)	1.33	31.02-36.22
No	121,664,683 (66.38)	1.33	63.78-68.98
Dependent variables			
Achieve health goal with tablet or smartphone			
Yes	66,165,939 (36.1)	1.60	32.96-39.25
No	117,119,211 (63.9)	1.60	60.75-67.04
Sent or received an SMS text message from the doctor			
Yes	55,022,202 (30.02)	1.40	27.27-32.78
No	128,262,948 (69.98)	1.40	67.22-72.73

Use of Tablet or Smartphone to Achieve Health-Related Goals

In the full model predicting *achieving health-related goals with the help of tablet or smartphone*, age, gender, marital status, ownership of basic cellphones, having a health-related wellness app, making health treatment decisions with mHealth, using devices other than tablets or smartphones to monitor or track health, and having a discussion with health care provider with the help of tablet or smartphone were significant predictors (Figure 1). In terms of the impact on the odds of achieving health-related goals with the help of tablet or smartphone, increasing age decreased the odds (35-49 years, odds ratio [OR] 0.41, 95% CI 0.18-0.91; 50-64 years, OR 0.17, 95% CI 0.08-0.38; 65-74 years, OR 0.11, 95% CI 0.04-0.29; >75 years, OR 0.07, 95% CI 0.02-0.19), being a woman increased the odds (OR 1.69, 95% CI 1.06-2.68), being married (OR 2.28, 95% CI 1.17-4.47) or previously married (OR 2.39, 95% CI 1.09-5.25) increased the odds, having a basic cellphone (OR 0.43, 95% CI 0.21-0.87) decreased the odds, having a wellness app (OR 8.70, 95% CI 5.81-13.04) increased the odds, making health decisions with mHealth (OR 1.77, 95% CI 1.06-2.94) increased the odds, tracking health with other devices (OR 2.73, 95% CI 1.46-5.12) and having discussion with the provider (OR 1.96, 95% CI 1.22-3.17) using tablet or smartphone increased the odds. Age, female gender, being married or previously married were also significant predictors of achieving health-related goals with the help of a tablet or smartphone when we accounted for only demographic variables (Multimedia Appendix 1). The reference categories for categorical predictors showed in Figure 1 include: age, 18-34 years; gender, male; education level, less than high school; race or ethnicity, non-Hispanic White; marital status, never married; household income, ≥US $75,000; employment status, unemployed; smoking, yes response; health status, fair; and other variables, no response. Also note the full expansion of the abbreviated variables in the figure as follows: Shared health device information: shared health information from electronic devices, smartphones, or tablets with health care providers; Other device track health: used other devices apart from tablet or smartphone to track health; Discuss with HCP: discuss with health care providers with the help of tablets or smartphones.

Figure 1. Full model with design-adjusted estimates of odds ratios for achieving health-related goal with the help of a tablet or smartphone among the hypertensive population (*P* values: [a].001, [b].01, [c].05). HCP: health care provider; mHealth: mobile health.



Use of Tablet or Smartphone to Communicate With Health Care Provider Through Text Messaging

In the full model predicting *send or receive SMS text messages to or from a health care professional in the last 12 months*, electronic communication with the doctor or doctor's office via email or internet (OR 2.93, 95% CI 1.85-4.63) and having health-related wellness apps (OR 1.82, 95% CI 1.16-2.86) were the only significant predictor variables (Table 3). Individuals who used a computer, smartphone, or other electronic means to use email or the internet to communicate with a doctor or doctor's office in the past 12 months had 193% higher odds of sending or receiving SMS text messages from a health care professional in the last 12 months than those who did not. Those with health-related wellness apps had 82% higher odds of sending or receiving text messages from a health care professional in the last 12 months than those who did not. No other covariates were found to be statistically significant.

Table 3. Full model with design-adjusted estimates of odds ratios for sending or receiving text message from health care provider in the last 12 months among the hypertensive population (sample size=3045; estimated population size=183,285,150).

Predictor and category	Odds ratio (95% CI)	SE	*P* value
Age group[a] (years)			
35-49	2.22 (0.75-6.58)	0.55	.17
50-64	1.56 (0.54-4.53)	0.55	.43
65-74	1.48 (0.45-4.84)	0.61	.53
≥75	1.55 (0.43-5.59)	0.66	.52
Gender[b]			
Female	1.20 (0.83-1.74)	0.19	.36
Education level[c]			
High school graduate	0.84 (0.37-1.87)	0.41	.67
Some college	0.68 (0.30-1.52)	0.42	.36
College graduate or more	0.61 (0.24-1.54)	0.47	.31
Race or ethnicity[d]			
Non-Hispanic Black or African American	0.75 (0.43-1.31)	0.29	.33
Hispanic	0.90 (0.47-1.74)	0.34	.77
Non-Hispanic Asian	0.67 (0.23-2.06)	0.57	.495
Non-Hispanic other	1.57 (0.37-6.74)	0.74	.55
Marital status[e]			
Married	1.37 (0.68-2.77)	0.36	.39
Previously married	1.54 (0.65-3.65)	0.44	.35
Household yearly income[f] (US $)			
<20,000	0.50 (0.22-1.15)	0.42	.13
20,000 to <35,000	0.55 (0.31-0.98)	0.29	.06
35,000 to <50,000	0.63 (0.34-1.16)	0.31	.16
50,000 to <75,000	0.82 (0.54-1.25)	0.22	.37
Employment status[g]			
Employed	0.80 (0.44-1.43)	0.30	.46
Smoked at least 100 cigarettes[h]			
No	1.21 (0.87-1.68)	0.17	.28
Health status[i]			
Very good	1.17 (0.66-2.09)	0.29	.60
Good	1.28 (0.79-2.09)	0.25	.33
BMI	1.00 (0.96-1.03)	0.02	.78
Diabetes[j]			
Yes	1.05 (0.69-1.60)	0.21	.82
Heart condition[j]			
Yes	1.04 (0.66-1.63)	0.23	.88
Depression[j]			
Yes	1.19 (0.70-2.05)	0.28	.53
Have smartphone[j]			

Predictor and category	Odds ratio (95% CI)	SE	*P* value
Yes	1.53 (0.62-3.80)	0.46	.37
Have tablet[j]			
Yes	1.08 (0.68-1.74)	0.24	.75
Have basic cellphone[j]			
Yes	0.79 (0.35-1.77)	0.41	.57
Have health apps[j]			
Yes	1.82 (1.16-2.86)	0.23	.02
Make treatment decision with mobile health[j]			
Yes	1.31 (0.84-2.04)	0.23	.26
Discuss with health provider with help of tablet or smartphone[j]			
Yes	0.98 (0.64-1.51)	0.22	.94
Used other devices apart from tablet and smartphone to monitor or track health[j]			
Yes	1.20 (0.79-1.84)	0.22	.40
Shared health info from electronic device, tablet, or smartphone with health provider[j]			
Yes	1.62 (1.02-2.56)	0.23	.06
Not applicable	0.74 (0.22-2.49)	0.62	.64
Electronic communication with doctor or doctor's office via email or internet[j]			
Yes	2.93 (1.85-4.63)	0.23	<.001

[a]Reference category: 18 to 34 years.
[b]Reference category: Male.
[c]Reference category: Less than high school.
[d]Reference category: Non-Hispanic White.
[e]Reference category: Never married.
[f]Reference category: ≥US $75,000.
[g]Reference category: Unemployed.
[h]Reference category: Yes response.
[i]Reference category: Fair.
[j]Reference category: No response.

Notably, in the model with only demographics, annual household income was the only significant predictor of sending or receiving SMS text messages to or from a health care professional in the last 12 months (Multimedia Appendix 2). Compared with the subpopulation with yearly household income of ≥US $75,000, the odds of sending or receiving SMS text messages from health care professionals in the last 12 months decreased by 40.0%, 50.7%, 64.7%, and 74.0%, respectively, among those with US $50,000 to <US $75,000 (OR 0.60, 95% CI 0.41-0.87; *P*=.01); US $35,000 to <US $50,000 (OR 0.49, 95% CI 0.28-0.88; *P*=.03); US $20,000 to <US $35,000 (OR 0.35, 95% CI 0.20-0.61; *P*=.001); and <US $20,000 (OR 0.26, 95% CI 0.13-0.51; *P*<.001) household incomes. The design-adjusted Wald test indicated that household income remained a significant predictor of sending or receiving SMS text messages to or from a health care professional ($F_{4,23}$=4.92, *P*=.005).

Discussion

Principal Findings and Implications

The purpose of this study was to identify predictors of *using a smartphone or tablet to achieve health goals* and SMS *text messaging communication with health care professionals* among individuals with hypertension. Most of the hypertensive population have a smartphone, and just over half have tablets. We found that the likelihood of using a smartphone or tablet to achieve health-related goals significantly decreased with increase in age and ownership of a basic cellphone. The use of smartphones or tablets to achieve health-related goals was, however, statistically significantly positively associated with being a woman, being married or previously married, having a health-related wellness app, making health treatment decisions with mHealth, using devices other than tablets or smartphones to monitor or track health, and having a discussion with a health care provider with the help of a tablet or smartphone. Sending or receiving SMS text messages to or from a health care provider

was statistically significantly positively associated with previous electronic communication with the doctor or doctor's office by email or internet and having a health-related wellness app.

Achieving health-related goals with the help of a tablet or smartphone usually involves having a health-related app installed on a smartphone or tablet [29]. Therefore, it is not surprising that age was a significant predictor of achieving health-related goals with the help of a tablet or smartphone, with the odds decreasing as age increases. This may be because younger people are more likely to have smartphones, tablets, and health-related apps [10,18]. Studies have shown that older adults can use technology if they understand the benefits that they can get from such use [30-32]. Health care providers can recommend that older patients use their smartphones and tablets to achieve health goals and encourage more use. Our findings among people with hypertension in younger age and female gender as significant predictors of achieving health-related goals with the help of a tablet or smartphone agree with another HINTS study [19] among respondents with one or more chronic diseases, which found that respondents aged 65 years had lower odds compared with those aged between 18 and 34 years, whereas women had higher odds compared with men for tracking the progress of health-related goals with their tablet or smartphone. They also found that those with health-related apps have higher odds of tracking the progress of health-related goals with their tablet or smartphone than those who do not have the app, which agrees with our findings as well. However, their findings showed that being employed increases the odds and having good health status decreases the odds of tracking the progress of health-related goals with tablets or smartphones, which differ from ours, where employment and health status were not associated with tracking health goals. The difference in results could be because of the differences in the variables included in the regression models, or it could also be because their study was conducted among respondents with one or more chronic diseases. Another HINTS study [20] on adult respondents also found that the likelihood of achieving health goals with the help of the mHealth app decreased with increasing age.

It was not surprising that those who own only basic cellphones had lower odds of achieving health-related goals with the help of a tablet or smartphone compared with those who did not, because basic cellphones are not usually equipped with advanced features to do that. The significance of making health treatment decisions with mHealth, using devices other than a tablet or smartphone to monitor or track health and having a discussion with health care providers with the help of a tablet or smartphone as predictors for achieving health-related goals with the help of a tablet or smartphone buttresses the fact that people who are already using a technology device are more likely to increase their use than those who are not using a technology device. This finding suggests that these groups of the hypertensive population can also benefit from telemonitoring of BP. Health care providers can play a role in creating awareness of these resources and their usefulness among their patients. It is also critical that payment reform adequately recognizes providers' time in supporting the telemonitoring of BP. The number of patients with hypertension using any of these technology devices could be increased by making them more affordable and accessible, and insurance coverage of such technology is likely necessary for the widespread adoption of telemonitoring of BP.

Less than one-third of the hypertensive subpopulation sent or received SMS text messages from their health care professionals. Interestingly, annual household income was a significant predictor of sending or receiving SMS text messages from a health care professional while considering only demographic variables, with the odds of sending or receiving SMS text messages decreasing with lower household income. SMS text messaging has been portrayed as a low-cost and common resource that can be used to improve health care. It has been shown to be effective in several intervention studies to improve hypertension knowledge and behavior changes, such as medication adherence and BP monitoring, leading to better BP control [13-16]. In general, 2-way SMS text messaging communication initiated by the health care provider keeps the patient and provider in frequent communication and is more effective in BP target attainment [14]. Our results show that the advantages of SMS text messaging are not being fully used in everyday life. One would think that SMS text messaging will be widespread across all income levels as it is considered an inexpensive option, but that is not the case. Advocacy for free SMS text messaging phone subscriptions for lower income patients with hypertension may increase the use of this technology. It could also be that patients are not aware that they can communicate with their health care professionals through SMS text messaging or that the service is not offered by their health providers. With adequate reimbursement, it behooves health care providers to initiate SMS text messaging with their patients so that they can both reap its advantages and free office time, to some extent, for more acute or serious visits.

In the full model, electronic communication with the doctor or doctor's office and having a health-related wellness app were significant predictors of sending or receiving SMS text messages from a health care professional when we accounted for all covariates. This shows that those already in communication with their health care providers are more likely to continue even if there is a change in the communication channel. The significance of having a health-related wellness app suggests that those who are already doing a form of self-monitoring are more likely to communicate with their health care provider via SMS text messaging. These findings are important because the impact of demographic characteristics such as age, gender, and income was not statistically significant. This suggests that all individuals, not just the young or those with higher incomes, for example, should be targeted to use remote BP monitoring. An essential first step may be the first electronic communication with the doctor or doctor's office, including email, electronic health portal messaging, or phone SMS text messaging. Having a health-related wellness app predicted both using a tablet or smartphone to achieve health goals and communication with a health care provider through SMS text messaging. These findings underscore the importance of these technology apps in improving health and the importance of the willingness of the patients to be more involved in their care through technology use. Health care systems could offer user-friendly health-related

wellness apps to patients on a secure platform to boost patient trust and increase uptake.

Study Limitations

Our study results are limited by the cross-sectional nature of the data, and the subpopulation used is based on self-reported hypertension. However, our robust analytic approach, which accounts for the HINTS sampling design, is positive. We also may not have accounted for all factors needed to predict the dependent variables because we used secondary data.

Future Studies

Effective engagement with health technology requires patients to have some eHealth literacy [33]. eHealth literacy expresses a person's understanding of the knowledge, skills, and resources needed to properly use health technology services. Future research should consider how factors, such as the eHealth literacy status of the patients or health care resources available to the patients are associated with the use of tablets or smartphones to achieve health goals and communicate with a health care provider through SMS text messaging in patients with hypertension. For example, mobile SMS text messages aimed at controlling child dental caries among parents with low eHealth literacy led to improvements in health outcomes and an increase in parental eHealth literacy in the intervention group at 6 months [34]. Further studies are required to understand how these predictors correlate with objective BP control and patient-provider communication preferences among patients with hypertension.

Conclusions

The use of mHealth to achieve health goals and communicate with health care professionals by patients with hypertension is significantly associated with having health-related wellness apps. Achieving health goals is also associated with demographics, such as age, gender, marital status, technology access, and other technology-related behaviors. Communication with health care providers through SMS text messaging is associated with previous electronic communications with the doctor or doctor's office. It is essential to consider these factors in tandem when planning telemonitoring for patients with hypertension. Measures accounting for these factors are required to increase smartphone and tablet use and their benefits in the routine care of patients with hypertension.

Authors' Contributions

KBF, MPD, and CEE conceived the project idea. CEE conducted the analysis and wrote the manuscript. BTW provided guidance on statistical analysis methods. BTW, MPD, ABC, CAL, LRB, and KBF reviewed the manuscript. All authors approved the final version of the manuscript.

Conflicts of Interest

Karen B Farris is the site principal investigator for an investigator-initiated grant from AstraZeneca examining coordination of care between oncology and primary care pharmacists. Dr Dorsch is supported by R18 HS026874 and R21 HS026322 from the Agency for Health Research and Quality, R01 AG062582 from the National Institutes of Health and National Institute of Aging, and the American Health Association Health IT Research Network; has received honoraria from Janssen; and has received research funding from Bristol Myers Squibb/Pfizer and Amgen in the past 2 years.

Multimedia Appendix 1
Demographics-only model with design-adjusted estimates of odds ratios for achieving health-related goals with the help of tablets or smartphones among the hypertensive population.
[DOCX File , 17 KB - jmir_v24i1e33188_app1.docx]

Multimedia Appendix 2
Demographics-only model with design-adjusted estimates of odds ratios for sending or receiving text messages from health care providers in the last 12 months among the hypertensive population.
[DOCX File , 17 KB - jmir_v24i1e33188_app2.docx]

References

1. Virani SS, Alonso A, Aparicio HJ, Benjamin EJ, Bittencourt MS, Callaway CW, American Heart Association Council on Epidemiology and Prevention Statistics Committee and Stroke Statistics Subcommittee. Heart Disease and Stroke Statistics-2021 update: a report from the American Heart Association. Circulation 2021 Feb 23;143(8):254-743 [FREE Full text] [doi: 10.1161/CIR.0000000000000950] [Medline: 33501848]
2. Uhlig K, Patel K, Ip S, Kitsios GD, Balk EM. Self-measured blood pressure monitoring in the management of hypertension: a systematic review and meta-analysis. Ann Intern Med 2013 Aug 06;159(3):185-194. [doi: 10.7326/0003-4819-159-3-201308060-00008] [Medline: 23922064]
3. Ishii M, Ogawa H, Unoki T, An Y, Iguchi M, Masunaga N, et al. Relationship of hypertension and systolic blood pressure with the risk of stroke or bleeding in patients with atrial fibrillation: the Fushimi AF registry. Am J Hypertens 2017 Nov 01;30(11):1073-1182. [doi: 10.1093/ajh/hpx094] [Medline: 28575205]

4. Kannan A, Janardhanan R. Hypertension as a risk factor for heart failure. Curr Hypertens Rep 2014 Jul;16(7):447. [doi: 10.1007/s11906-014-0447-7] [Medline: 24792121]
5. Yandrapalli S, Nabors C, Goyal A, Aronow WS, Frishman WH. Modifiable risk factors in young adults with first myocardial infarction. J Am Coll Cardiol 2019 Feb 12;73(5):573-584 [FREE Full text] [doi: 10.1016/j.jacc.2018.10.084] [Medline: 30732711]
6. Aronow WS. Treatment of systemic hypertension. Am J Cardiovasc Dis 2012;2(3):160-170 [FREE Full text] [Medline: 22937486]
7. Billups SJ, Saseen JJ, Vande Griend JP, Schilling LM. Blood pressure control rates measured in specialty vs primary care practices within a large integrated health system. J Clin Hypertens (Greenwich) 2018 Sep;20(9):1253-1259 [FREE Full text] [doi: 10.1111/jch.13345] [Medline: 30009553]
8. Chandak A, Joshi A. Self-management of hypertension using technology enabled interventions in primary care settings. Technol Health Care 2015;23(2):119-128. [doi: 10.3233/THC-140886] [Medline: 25515051]
9. Internet/broadband fact sheet. Pew Research Center. URL: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/ [accessed 2021-12-29]
10. Mobile fact sheet. Pew Research Center. 2021. URL: https://www.pewresearch.org/internet/fact-sheet/mobile/ [accessed 2021-12-29]
11. Omboni S, Caserini M, Coronetti C. Telemedicine and m-health in hypertension management: technologies, applications and clinical evidence. High Blood Press Cardiovasc Prev 2016 Sep;23(3):187-196. [doi: 10.1007/s40292-016-0143-6] [Medline: 27072129]
12. Margolis KL, Asche SE, Bergdall AR, Dehmer SP, Groen SE, Kadrmas HM, et al. Effect of home blood pressure telemonitoring and pharmacist management on blood pressure control: a cluster randomized clinical trial. J Am Med Assoc 2013 Jul 03;310(1):46-56 [FREE Full text] [doi: 10.1001/jama.2013.6549] [Medline: 23821088]
13. Hacking D, Haricharan HJ, Brittain K, Lau YK, Cassidy T, Heap M. Hypertension health promotion via text messaging at a community health center in South Africa: a mixed methods study. JMIR Mhealth Uhealth 2016 Mar 10;4(1):e22 [FREE Full text] [doi: 10.2196/mhealth.4569] [Medline: 26964505]
14. Vargas G, Cajita MI, Whitehouse E, Han H. Use of short messaging service for hypertension management: a systematic review. J Cardiovasc Nurs 2017;32(3):260-270 [FREE Full text] [doi: 10.1097/JCN.0000000000000336] [Medline: 27111819]
15. Buis L, Hirzel L, Dawood RM, Dawood KL, Nichols LP, Artinian NT, et al. Text messaging to improve hypertension medication adherence in African Americans from primary care and emergency department settings: results from two randomized feasibility studies. JMIR Mhealth Uhealth 2017 Mar 01;5(2):e9 [FREE Full text] [doi: 10.2196/mhealth.6630] [Medline: 28148474]
16. Zhai P, Hayat K, Ji W, Li Q, Shi L, Atif N, et al. Efficacy of text messaging and personal consultation by pharmacy students among adults with hypertension: randomized controlled trial. J Med Internet Res 2020 May 20;22(5):e16019 [FREE Full text] [doi: 10.2196/16019] [Medline: 32432556]
17. Duan Y, Xie Z, Dong F, Wu Z, Lin Z, Sun N, et al. Effectiveness of home blood pressure telemonitoring: a systematic review and meta-analysis of randomised controlled studies. J Hum Hypertens 2017 Jul;31(7):427-437. [doi: 10.1038/jhh.2016.99] [Medline: 28332506]
18. Langford AT, Solid CA, Scott E, Lad M, Maayan E, Williams SK, et al. Mobile phone ownership, health apps, and tablet use in US adults with a self-reported history of hypertension: cross-sectional study. JMIR Mhealth Uhealth 2019 Jan 14;7(1):e12228 [FREE Full text] [doi: 10.2196/12228] [Medline: 31344667]
19. Mahmood A, Kedia S, Wyant DK, Ahn S, Bhuyan SS. Use of mobile health applications for health-promoting behavior among individuals with chronic medical conditions. Digit Health 2019;5:2055207619882181 [FREE Full text] [doi: 10.1177/2055207619882181] [Medline: 31656632]
20. Bhuyan SS, Lu N, Chandak A, Kim H, Wyant D, Bhatt J, et al. Use of mobile health applications for health-seeking behavior among US adults. J Med Syst 2016 Jun;40(6):153. [doi: 10.1007/s10916-016-0492-7] [Medline: 27147516]
21. Krebs P, Duncan DT. Health app use among US mobile phone owners: a national survey. JMIR Mhealth Uhealth 2015 Nov 04;3(4):e101 [FREE Full text] [doi: 10.2196/mhealth.4924] [Medline: 26537656]
22. Methodology reports - Health Information National Trends Survey 5 (HINTS 5) Cycle 1. National Cancer Institute. 2017. URL: https://hints.cancer.gov/data/methodology-reports.aspx [accessed 2021-12-29]
23. Methodology reports - Health Information National Trends Survey 5 (HINTS 5) Cycle 2. National Cancer Institute. 2018. URL: https://hints.cancer.gov/data/methodology-reports.aspx [accessed 2021-12-29]
24. Heeringa S, West B, Berglund P. Preparation for complex sample survey data analysis. In: Applied Survey Data Analysis. London, United Kingdom: Chapman and Hall; 2017.
25. Lumley T, Scott A. AIC and BIC for modeling with complex survey data. J Survey Stat Methodol 2015 Jan 30;3(1):1-18. [doi: 10.1093/jssam/smu021]
26. Echouffo-Tcheugui JB, Batty GD, Kivimäki M, Kengne AP. Risk models to predict hypertension: a systematic review. PLoS One 2013;8(7):e67370 [FREE Full text] [doi: 10.1371/journal.pone.0067370] [Medline: 23861760]

27. Ware LJ, Chidumwa G, Charlton K, Schutte AE, Kowal P. Predictors of hypertension awareness, treatment and control in South Africa: results from the WHO-SAGE population survey (Wave 2). J Hum Hypertens 2019 Feb;33(2):157-166. [doi: 10.1038/s41371-018-0125-3] [Medline: 30382179]
28. Acheampong K, Nyamari JM, Ganu D, Appiah S, Pan X, Kaminga A, et al. Predictors of hypertension among adult female population in Kpone-Katamanso District, Ghana. Int J Hypertens 2019;2019:1876060 [FREE Full text] [doi: 10.1155/2019/1876060] [Medline: 31308975]
29. Omboni S. Connected health in hypertension management. Front Cardiovasc Med 2019;6:76 [FREE Full text] [doi: 10.3389/fcvm.2019.00076] [Medline: 31263703]
30. Peek STM, Luijkx KG, Rijnaard MD, Nieboer ME, van der Voort CS, Aarts S, et al. Older adults' reasons for using technology while aging in place. Gerontology 2016;62(2):226-237 [FREE Full text] [doi: 10.1159/000430949] [Medline: 26044243]
31. Pruchno R. Technology and aging: an evolving partnership. Gerontologist 2019 Jan 09;59(1):1-5. [doi: 10.1093/geront/gny153] [Medline: 30629258]
32. Pelizaus-Hoffmeister H. Motives of the elderly for the use of technology in their daily lives. In: Ageing and Technology. Bielefeld: Transcript Verlag; 2016.
33. Norgaard O, Klokker BL, Hospital F, Astrid KD, Kayser L, Osborne R. The e-health literacy framework: a conceptual framework for characterizing e-health users and their interaction with e-health systems. Knowl Manag E-Learn Int J 2015;7(4):522-540. [doi: 10.34105/j.kmel.2015.07.035]
34. Lotto M, Strieder AP, Aguirre PE, Oliveira TM, Machado MA, Rios D, et al. Parental-oriented educational mobile messages to aid in the control of early childhood caries in low socioeconomic children: a randomized controlled trial. J Dent 2020 Oct;101:103456 [FREE Full text] [doi: 10.1016/j.jdent.2020.103456] [Medline: 32827598]

Abbreviations

BP: blood pressure
HINTS: Health Information National Trends Survey
mHealth: mobile health
OR: odds ratio

Edited by R Kukafka; submitted 31.08.21; peer-reviewed by K Trinkley, M Ofori, M Lotto, S Hajesmaeel Gohari; comments to author 22.10.21; revised version received 11.11.21; accepted 03.12.21; published 24.01.22.

Please cite as:
Eze CE, West BT, Dorsch MP, Coe AB, Lester CA, Buis LR, Farris K
Predictors of Smartphone and Tablet Use Among Patients With Hypertension: Secondary Analysis of Health Information National Trends Survey Data
J Med Internet Res 2022;24(1):e33188
URL: https://www.jmir.org/2022/1/e33188
doi:10.2196/33188
PMID:35072647

Original Paper

General Practitioners' Attitudes Toward Artificial Intelligence–Enabled Systems: Interview Study

Christoph Buck[1,2*], PhD; Eileen Doctor[3*], MSc; Jasmin Hennrich[3*], MSc; Jan Jöhnk[4*], PhD; Torsten Eymann[1,4*], Prof Dr

[1]Department of Business & Information Systems Engineering, University of Bayreuth, Bayreuth, Germany

[2]Centre for Future Enterprise, Queensland University of Technology, Brisbane, Australia

[3]Project Group Business & Information Systems Engineering, Fraunhofer Institute for Applied Information Technology, Bayreuth, Germany

[4]Finance & Information Management Research Center, Bayreuth, Germany

[*]all authors contributed equally

Corresponding Author:
Christoph Buck, PhD
Department of Business & Information Systems Engineering
University of Bayreuth
Universtitätsstraße 30
Bayreuth, 95447
Germany
Phone: 49 (0)921 557665
Email: Christoph.Buck@uni-bayreuth.de

Abstract

Background: General practitioners (GPs) care for a large number of patients with various diseases in very short timeframes under high uncertainty. Thus, systems enabled by artificial intelligence (AI) are promising and time-saving solutions that may increase the quality of care.

Objective: This study aims to understand GPs' attitudes toward AI-enabled systems in medical diagnosis.

Methods: We interviewed 18 GPs from Germany between March 2020 and May 2020 to identify determinants of GPs' attitudes toward AI-based systems in diagnosis. By analyzing the interview transcripts, we identified 307 open codes, which we then further structured to derive relevant attitude determinants.

Results: We merged the open codes into 21 concepts and finally into five categories: concerns, expectations, environmental influences, individual characteristics, and minimum requirements of AI-enabled systems. Concerns included all doubts and fears of the participants regarding AI-enabled systems. Expectations reflected GPs' thoughts and beliefs about expected benefits and limitations of AI-enabled systems in terms of GP care. Environmental influences included influences resulting from an evolving working environment, key stakeholders' perspectives and opinions, the available information technology hardware and software resources, and the media environment. Individual characteristics were determinants that describe a physician as a person, including character traits, demographic characteristics, and knowledge. In addition, the interviews also revealed the minimum requirements of AI-enabled systems, which were preconditions that must be met for GPs to contemplate using AI-enabled systems. Moreover, we identified relationships among these categories, which we conflate in our proposed model.

Conclusions: This study provides a thorough understanding of the perspective of future users of AI-enabled systems in primary care and lays the foundation for successful market penetration. We contribute to the research stream of analyzing and designing AI-enabled systems and the literature on attitudes toward technology and practice by fostering the understanding of GPs and their attitudes toward such systems. Our findings provide relevant information to technology developers, policymakers, and stakeholder institutions of GP care.

(J Med Internet Res 2022;24(1):e28916) doi:10.2196/28916

KEYWORDS

artificial intelligence; AI; attitude; primary care; general practitioner; GP; qualitative interview; diagnosis; clinical decision support system

Introduction

Overview

As artificial intelligence (AI) enabled systems have surpassed human performance in different aspects of economy and society, the increasing technological maturity and widespread applicability of such systems is leading to skyrocketing expectations [1]. The technological progress in various fields such as machine learning, robotics, big data analytics, decision support systems (DSSs) as well as the ubiquity and availability of data and the prevalence of information systems (ISs) are opening previously unavailable value creation potentials [2-5]. We understand AI as a set of value-adding technological solutions that use self-learning algorithms to perform cognitive tasks at a level comparable with that of humans [6]. Various AI solutions that provide decision support typically associated with human cognition are emerging and hold the potential to reshape the nature of work [1,7-9]. Thus, AI is also a promising approach for the health care domain [10]. AI and related technologies, such as big data analytics and DSSs, are distinct phenomena with important conceptual differences; however, some of the underlying technologies might overlap. In health care, AI technology advances health information technologies such as clinical DSSs (CDSSs). These systems assist medical professionals in tasks related to medical decision-making [11], such as diagnosis, prescription, or the prevention of medication errors [12,13]. Among others, typical functions are alerts, reminders, and recommendations [14,15].

There are two forms of CDSSs in health care: knowledge-based systems and non–knowledge-based systems. Knowledge-based systems match their knowledge base with individual patient characteristics and make decisions based on preformulated rules [16,17]. As such, knowledge-based CDSSs are designed to inform skilled actors. That is, to provide actors in the health care system, for example, physicians, with relevant information to comprehend internal and external structures and processes. On the other hand, non–knowledge-based CDSSs use AI technologies, which eliminate the writing of rules and the need to follow expert medical input. This integration of AI technology allows the CDSSs to learn from experience and find patterns in medical data [18]. Hence, the vision of AI is to enable systems to be on human-level intelligence. Here, intelligence refers to an agent's ability to achieve goals in a wide range of environments [19,20] and goes beyond the mere preparation of information. Instead, AI highlights the ambition to develop artificial agents that are able to learn, decide, and act autonomously [9,21].

AI-enabled systems have already successfully entered various subdisciplines of health care, such as image recognition, diagnosis, and precision medicine [10]. Most AI-enabled systems have immediate relevance in health care and several potentials for value creation, such as higher efficiency and accuracy in diagnosis and lower error rates [22-24]. Furthermore, AI-enabled systems are more enduring in repetitive tasks than humans, thus enhancing cost-efficiency [25].

Regarding these promised benefits, AI-enabled systems have particular potential in the field of primary care. General practitioners (GPs) serve as the first point of medical contact and therefore must diagnose with high levels of uncertainty and under high time pressure. For instance, in Germany, primary care is one of the most frequently used health care services, leading to an average physician–patient contact time of 7.6 minutes [26]. Moreover, GPs are responsible for the initial diagnosis, thus setting the direction for whether a patient receives the right care. Misdiagnosis in this early stage of diseases can have severe impact on medical quality in terms of injuries, avoidable illnesses, hospitalizations, and in 10% of cases, death [27,28]. Besides the potentially tragic individual consequences, such misdiagnoses also increase the cost of care [29].

To prevent these risks, the health care system depends on innovative, reliant, and fast approaches to decision-making processes in GP care [30]. Considered as an integrative system, AI-enabled systems free up physicians' time for more sophisticated tasks [31]. Furthermore, AI-enabled systems can ensure stronger physician–patient relationships [32], which is especially valuable in GP care as it enables the therapeutic benefit of improved continuity of care and more holistic and individualized treatments [32]. In addition, AI-enabled systems can reduce diagnostic errors, which are considered the greatest threat to patient safety in GP care [33].

Although AI-enabled systems in primary care diagnosis are gradually becoming feasible and useful, their widespread implementation still remains a future scenario [10,34]. Among others, reasons for the slowdown in adoption are the physicians' lack of trust in [35,36] and acceptance of [16,37] AI-enabled systems. These adoption barriers arise, for instance, from the concern that AI-enabled systems might be trained with a heterogeneous database owing to the diversity and individuality of medicine, leading to biased or overadapted outcomes. Overcoming these hurdles requires balancing the GP's trust in AI-enabled systems [35]. On one hand, developing trust in such a system is beneficial to its adoption and use. On the other hand, AI-enabled systems may bear risks when physicians blindly rely on such systems' suggestions and outcomes. Furthermore, factors such as the anticipated threat to professional autonomy and legal liabilities from using AI-enabled systems are hindering factors, as known so far [38].

A key driver for successful implementation and uptake of AI-enabled systems is the attitude of physicians. By discussing our findings on GPs' attitudes toward AI-enabled systems within a facilitating context for practical implementation, we extend the previous work of Blease et al [39], who recognized the relevance of the topic and investigated the opinions of GPs about the possible impact of AI on GP care.

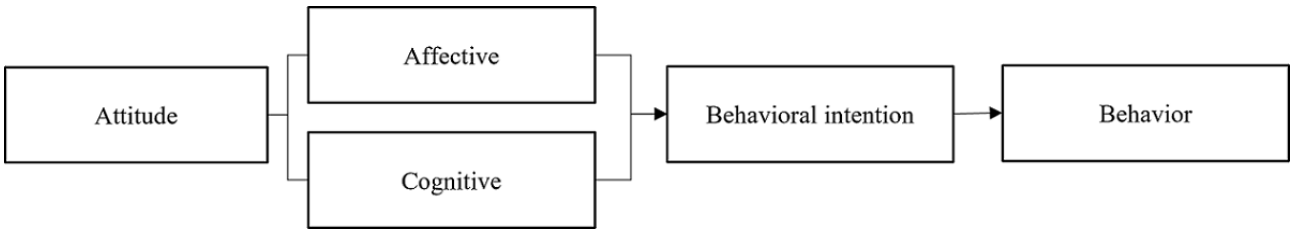
The Construct of Attitude as Our Theoretical Lens

We understand attitude as a psychological tendency that determines how GPs evaluate their favor or disfavor against AI-enabled systems [40]. Following Rosenberg and Hovland [41], the most widespread construct of attitude—the 3-component model—comprises the affective, cognitive, and behavioral dimensions of attitude. First, the affective component refers to the respondent's emotional reaction to an attitude object, including their empathy, preferences, and feelings.

Second, a person's thoughts and beliefs toward an attitude object form the cognitive component, which includes the individual's idea, opinion, or knowledge of it [41]. Third, the behavioral component rests on the attitude behavioral consistency assumption, described as the extent to which an attitude predicts a behavior, including the willingness or intention to act to deal with an object [41,42]. Overall, the attitude construct assumes a consistent and dependent relationship among the affective, cognitive, and behavioral components, suggesting that a change in one component leads to changes in the other components [41].

However, researchers acknowledge that behavioral intention (ie, the behavioral component) does not always correspond to the feelings (ie, the affective component) and opinions (ie, the cognitive component) [42]. This challenges the behavioral component as an integral part of the attitude construct. Thus, the 2-component model of attitude was developed based on this critique (Figure 1) [43]. According to this model, attitude consists of an affective and a cognitive component that simultaneously form the behavioral intention, which—in turn—explains a de facto behavior [44].

Figure 1. The 2-component model of attitude [44,45].



Extant work describes behavioral intention as the mediator in the relationship between attitude and behavior [46,47]. Thus, it is assumed that the stronger the intention, the higher the likelihood of the behavior occurring [48]. According to the principle of compatibility, behavior is only predicted by attitude to the extent of both being on the same level of specificity or generality regarding their objective, context, and time elements [49]. However, regardless of the intensity of influence, there is broad agreement that attitude fosters behavioral intention [50]. Drawing on its relevance to users' subsequent intentions and behavior, we use the attitude construct to foster our understanding of how to better exploit AI-enabled systems' potential in GP care and promote their future use.

Different quantitative studies have investigated the relationship between attitude and the intention to use AI-enabled systems, for example, regarding medical students [51,52]. In this context, research theories such as the unified theory of acceptance and use of technology (UTAUT) and the theory of reasoned action find application. These established approaches of technology acceptance research, originating in social psychology, primarily focus on users' intentions [48,53-55]. However, these approaches do not provide a comprehensive understanding of GPs' attitudes toward AI-enabled systems, as they use abstract constructs and variables and do not capture detailed, even emotionally based, and spontaneous responses from the potential users [56]. However, apart from this, there is a more important criterion why we chose not to use these models for our study. GPs rarely use AI-enabled systems so far [34], so research in this respect considers a rather hypothetical use scenario than actual use. Investigating the intention to use, which is a direct determinant of the actual use according to UTAUT, is therefore not feasible as the possible system features and functions are not yet available. Nevertheless, it is possible to look at the underlying attitude toward the technology, which exists outside the de facto experience of use.

Goal of This Study

Despite the relevance of AI technologies in the health care sector, a profound understanding of GPs' attitudes toward AI-enabled systems and their underlying determinants is still lacking. Therefore, the purpose of this study is to investigate which determinants influence GPs' attitudes toward AI-enabled systems in diagnosis. We see this as an important step in developing user-centered solutions, which will positively affect the intention to use and support the successful introduction of AI-enabled systems in primary care.

Methods

Data Collection

To identify the determinants of GPs' attitudes toward AI-enabled systems in diagnosis, in-depth insights are vital. Following the interpretative paradigm, qualitative methods were used to obtain an understanding of individuals' technological attitudes in the medical context [57]. Thus, we did not prescribe and narrow the phenomenon to only the testing of variables but emphasized the complexity of human understanding and behavior [58].

Data collection followed an interplay of continuous and iterative matching steps of sample selection (recruiting of participants), interview guideline creation (and improvement), data collection (interview conduction), data analysis (transcription and coding), and revision of the process steps. As the iterative process and constant comparison make it challenging to provide a timeline or sequence of these steps, it is reflected upon as a constant effort in creating a comprehensive and growing understanding of the participants' attitudes, which are not always distinctively observable [59,60]. In terms of saturation approaches, this study emphasized the term *conceptual depth* proposed by Nelson [61], whereby researchers cumulatively judge the sufficiency of depth of understanding, thus allowing for incremental development. Following Schultze and Avital [60], the choice of semistructured expert interviews allowed focusing on the research topic while also providing in-depth information [62]. This approach offered a modular structure through which the participants could access and reflect upon their experiences and perceptions regarding AI-enabled systems. We derived overarching interview topics from the given practical research

objective and through reflective discussions within the author team, resulting in the exploratory interview questions. The 4-phase process to interview protocol refinement, proposed by Castillo-Montoya [63], served as a basis for developing the interview guideline, including pretesting the first version of the interview guideline with three volunteers: a health economist, a nurse, and a physician. The first version of the interview guide addressed the topics of personal experiences, assessments of perceived diagnostic support, design requirements, and motivations for the use of AI-enabled systems in a broader perspective. It underwent 9 iterations, receiving more detailed and tailored questions about the research topic with each interview. The final interview guideline (Multimedia Appendix 1) was designed to question the participants on their understanding of AI-enabled systems and provide this paper's literature-based definition, aiming toward a shared understanding of AI-enabled systems and comparable interview results. Owing to the nature of the semistructured interviews, the results were not limited to collecting attitude determinants. In addition, this approach allowed us to capture insights into GP care's challenges and special characteristics, which contributed to a profound understanding of the determinants of attitudes in the medical context.

The interviews took place both in person and via phone and anonymity was guaranteed to all participants within the study. As face-to-face interviews create a trusting and comfortable atmosphere and enable more detailed information on participants' feelings and attitudes, the interviewers preferred them for data collection [64]. With the participants' consent, audio recording and transcription was performed to allow thorough data analysis by using a software program for qualitative and mixed methods data analysis, named *MAXQDA 2020* (VERBI GmbH).

Data Analysis

For analyzing the interview transcripts, grounded theory analysis techniques were applied. As stated by Glaser [65], the traditional grounded theory methodology (GTM) seeks to develop a conceptual theory that depicts a relevant or problematic behavior pattern (here, GPs' attitudes toward AI-enabled systems in diagnosis). GTM focuses on behavioral aspects where attitude behaves as an antecedent and is therefore equally suitable for application. Applying the GTM approach allowed to handle the unstructured qualitative data sets, discover relevant categories and relationships among them, and contextualize and interpret them [66]. According to GTM, the analysis begins with the first collected data set, as the experiences with the first interview process already influence the researcher and thus the upcoming interviews. In the interview process, participants' responses and clarified check-backs were closely scrutinized and documented [67]. This knowledge about misconceptions was considered in the iterative development of the interview guidelines. Furthermore, it allowed the clarification and precise alignment of the research question [66].

The interview data were paraphrased into relevant bits (open coding) in line with the 3-step Straussian approach for coding (open, axial, and selective coding). Thus, the first step consisted of an initial and careful reading of the interview transcripts, highlighting any phrases that may have proven to be relevant to the research topic. Over the course of data analysis, 307 open codes emerged. Following the Glaser and Strauss [68] specifications, the codes were further examined and paraphrased, merging those with common themes into concepts. Thus, we assigned special value to the wording of and syntactical differentiation among expressions. After comparing the allocation of the concepts, they were merged into categories. Moreover, the relationships among them were identified, which refers to the axial coding step. By setting all elements in relation to one another, the core category *attitude determinants* was distinguished from other categories (selective coding) [66]. In line with GTM, the 3 coding steps followed a flexible and iterative process instead of a fixed sequence [68].

For enhanced validity of the coding results, 2 authors (JH and ED) performed card-sorting allocation. Thus, the open codes and concepts identified by 1 author (JH) in the first round served as the foundation for the second author (ED) but in an unmatched format. The second author conducted a blind card-sorting round with this groundwork and commented on the constructs and documented challenges that arose in the allocation of open codes to a specific construct. This second author further added open codes that were not initially identified during the process. In case of deviations in matching open code to constructs between the 2 authors, the entire research team discussed the said allocations. An agreement was found in all cases of card-sorting deviations. Furthermore, in all coding rounds, the authors iteratively discussed the constructs' abstraction levels and their various definitions and revisited their coding results for adjustments, which the literature refers to as *constant comparison method* [66]. Whenever the authors gained new insights from their constant comparison and iterations, they repeated the open coding steps for all the interview sets backward and forward.

Results

Descriptive Results and Study Population

We interviewed 18 GPs from Germany between March 2020 and May 2020, selecting them via convenience sampling [69]. Thereby, we contacted 110 physicians within the geographic reach of the research team via mail and further relied on personal network contacts. In addition, we asked the acquired participants for the contact information of other colleagues who might be interested in participation. All participants had at least 1 year of work experience in GP care [69]. Of the 18 GPs, 7 (39%) were situated in urban areas with a range of 75,000-127,000 inhabitants, whereas 11 (61%) participants were situated in rural and small-town areas with a range of 3200-23,000 inhabitants. For a more accurate evaluation of the participants' statements in light of relevant demographic and structural data, individual characteristics of the participating GPs and descriptive characteristics of our data collection are shown in Table 1. We further report the specifics of the interview lengths and styles.

Table 1. Descriptive characteristics of the participants and the data collection (N=18).

Participant number	Age (years)[a]	Gender	Working situation	Interview duration (minutes)[b]	Interview style
GP[c] 1	70	Female	JP[d]	27	In person
GP 2	51	Male	JP	28	In person
GP 3	50	Male	JP	31	Via phone
GP 4	41	Male	JP	22	In person
GP 5	52	Female	JP	36	In person
GP 6	50	Female	__[e]	—	—
GP 7	50	Female	JP	23	Via phone
GP 8	36	Female	JP	25	In person
GP 9	45	Female	JP	23	Via phone
GP 10	58	Male	JP	46	In person
GP 11	38	Male	IP[f]	30	Via phone
GP 12	44	Female	JP	35	Via phone
GP 13	52	Male	JP	60	Via phone
GP 14	43	Female	JP	25	Via phone
GP 15	40	Male	GPC[g]	29	Via phone
GP 16	34	Female	IP	40	In person
GP 17	47	Male	JP	23	Via phone
GP 18	51	Male	IP	44	Via phone

[a]Mean age is 47.33 (SD 8.31) years.

[b]Mean interview duration is 30.38 (SD 10.06) minutes.

[c]GP: general practitioner.

[d]JP: joint practice.

[e]GP 5 and GP 6 participated in the interview together.

[f]IP: individual practice.

[g]GPC: general practitioner center.

Three-Step Coding Results

Overview

We describe the 5 categories and 21 concepts that determined our GPs' attitudes toward AI-enabled systems as derived from our qualitative data sets. Our baseline for considering the attitude determinants was the AI literacy level among the participants. Long and Magerko [70] defined AI literacy "as a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool on the web, at home, and in the workplace." Hence, the identified attitude determinants depend on the participants' statements and their knowledge regarding AI-enabled systems, irrespective of whether this knowledge is true to facts. Most participants had poor AI literacy in the data set and had not yet interacted with AI-enabled systems. For example, the self-learning ability of AI-enabled systems was known to only 33% (6/18) of the respondents. Although these 6 GPs were familiar with this AI technology component, they often did not fully understand what AI is. For example, GP 3 mentioned the following:

> *In the end, every time I turn on a computer, I use artificial intelligence.* [Participant 3]

Only 22% (4/18) of the GPs had experience with AI-enabled systems and only 50% (2/4) of them explicitly mentioned having used it in their GP work. In answering the question of why GPs had not had experiences with AI-enabled systems, the participants gave 3 explanations. First, they said they did not know about any AI-enabled tools for the GP sector (interview 15). Second, they did not see the necessity to use AI-enabled systems (interview 9). Third, is a general aversion toward the use of technology in medicine (interview 8). Although most participants had not had contact with AI-enabled systems, most GPs agreed on the role of the AI-enabled system in GP care in the future. A participant said the following:

> *You cannot decide against [AI technology] because it will come. Because without [AI technology] [diagnosis] is not possible.* [Participant 1]

The participants associated expected time effort with the use of AI-enabled systems in routine diagnoses owing to the necessary AI technology integration into an established and effortless routine process. Therefore, the participants limited the scope of its application to cases of rare diseases and to cases in which

the physicians could not reach a diagnosis without additional help (interview 8).

When grouping the statements, we paid particular attention to the wording and syntactic differentiation that the physicians used in their answers. The interview data revealed 5 main categories that summarize the influencing determinants of GPs' attitudes toward AI-enabled systems in diagnosis. When we raised questions on potentially using AI-enabled systems in clinical practice, the GPs had various *concerns* and *expectations*. In addition, we found that the *environmental influences* and certain *individual characteristics* influenced their attitudes. Whenever GPs stated that AI-enabled systems must meet certain requirements for them to consider using it, we categorized them as *minimum requirements of AI-enabled systems*. Table 2 shows an overview of all the categories and concepts, which is followed by a description of the determinants, as supported by interview quotes.

Table 2. Overview of the categories and concepts.

Determinants of attitudes toward AI[a]-enabled systems and concepts	Open codes in each concept	Open codes in each category
Concerns		57
Existential anxiety	12	
Change of the physician–patient relationship	7	
Misuse of data	14	
Diagnostic bias	24	
Expectations		112
Diagnostic quality	35	
Diagnostic efficiency	19	
Legal liability	4	
Lack of human competencies	43	
Time expenditure	11	
Environmental influences		37
Changing working conditions	8	
Stakeholder influences	13	
Media	12	
Information technology infrastructure	4	
Individual characteristics		17
Age	11	
Affinity with technology	6	
Minimum requirements of AI-enabled systems		84
Time efficiency	40	
Diagnostic quality	15	
Data security	10	
Economic viability	12	
Transparency	3	
Autonomy	4	

[a]AI: artificial intelligence.

Concerns

Overview

Concerns include all doubts and fears concerning AI-enabled systems. Overall, this category consists of four concepts: (1) existential anxiety, (2) change of the physician–patient relationship, (3) misuse of data, and (4) diagnostic bias.

Existential Anxiety

Half (9/18, 50%) of the participants expressed *existential anxiety* connected with AI-enabled systems as they perceive that this technology can take over some of their tasks. GP 2 said the following:

> *At one point, the own decision and the own expertise threatens to be pushed into the background or to become redundant.* [Participant 2]

GP 14 also perceived the threat of being replaceable by AI-enabled systems and provided an example of an AI-enabled system that has achieved higher diagnostic accuracy than physicians. This concept included the fear of no longer being useful and being replaceable by AI-enabled systems and the worry of losing their unique status as physicians. A participant said the following:

> *Surely, many doctors probably see their unique medical status endangered, that they are under the surveillance of others, that they think there is a bit of an attack on their own vanity.* [Participant 12]

Change of the Physician–Patient Relationship

The participants mentioned that AI-enabled systems could threaten the physician–patient relationship. Endangerment of this relationship, which fundamentally defines GP care, further compromises appropriate patient care (participant 3). As patients could feel that the AI-enabled system performs the treatment, the physicians assumed that the use of AI-enabled systems might negatively impact the physician–patient relationship (participant 11). In this regard, a participant mentioned the following:

> *Since [the patient] has the feeling [...] that the machine takes care of it and the doctor would only have to put his signature under it.* [Participant 11]

Participants mentioned the impairment of the physician–patient conversation through the use of technology as threatening to the physician–patient relationship. The concern is that, by using AI-enabled systems during patient consultations, a GP cannot devote all their attention to the patient sitting in front of them, but instead must also focus on the screen to follow an AI-enabled system's recommendations. A participant commented the following:

> *[The treatment] may drift off into a standardized interview, and that's probably not necessary.* [Participant 12]

GP 13 was concerned that AI-enabled systems would generally reduce physician–patient contact, which is a core component of GP care and is inevitable for successful treatment and patient care. The potential endangerment of the physician–patient relationship by the use of AI-enabled systems was also often linked to misuse of data.

Misuse of Data

With the use of AI-enabled systems and the disclosure of both the patients' and physicians' data, *misuse of data* is a key concern and impacts GPs' attitudes toward AI-enabled systems. In this context, GP 3 saw the problem in the connection between AI-enabled systems used in practice and the interconnectedness between these systems and the internet:

> *[AI-enabled systems] are not stand-alone systems but are networked, and [...] actually, work over the internet with such simple things as voice recognition. And in my view, this will change the doctor-patient contact considerably. [...] I consider the fundamental trust in the patient-physician-conversation [...] to be a very important basis for our work. And I also see [the trusting relationship between the patient and the physician] as being in danger due to the increasing use of such procedures. I find this very worrying.* [Participant 3]

Internet access makes the data accessible and renders the patient and physician transparent, thus violating data privacy and leading to serious consequences for patients. A participant described it as follows:

> *Patient data are very sensitive data. Disease data are very sensitive data. [There is the risk that] they are passed on somewhere, that some authorities who have nothing to do with it or should have nothing to do with it could intercept the data and use this to the disadvantage of the patients.* [Participant 11]

Thus, the physicians are concerned about the data being misused by other stakeholders as supported by GP 4:

> *The problem is that large companies use AI to gain access to lucrative patients and to control them via AI.* [Participant 4]

Further, GP 3 warned of the danger of pharmaceutical companies programming AI-enabled systems for their purposes, referring to medication proposals that are not medically indicated but instead deliver a monetary benefit for the producing company. They justified this concern with experiences from working with other technologies (participant 3). This concept also summarizes physicians' concerns about being monitorable and controllable at work when using AI-enabled systems. Owing to connection to the internet, GP 10 assumed that every step of physicians will be transparent and can be monitored. However, the GPs did not explicitly mention who would have interests in observing and controlling them.

Diagnostic Bias

According to the participants, AI-enabled systems can cause *diagnostic bias*, whereby the technology influences the GP's decision-making in ways that can negatively affect the course and success of treatment. Once a GP has received suggestions from an AI-enabled system, they may not consider further possible diagnoses (participant 11). In this context, GP 8 spoke of the fear of being put on a completely wrong track and the likelihood that the AI-enabled system indicates a diagnosis that does not fit and therefore leads a GP to mistreat the patient. A frequent concern was that physician might become overreliant on the technology, neglecting their own medical and experience-based knowledge. Furthermore, the participants also mentioned the risk of overexpansion of treatment services as supported by participant 17:

> *The AI will recommend examinations that I would personally put last, ie. it will possibly lead to so-called device medicine, involving a lot of safeguard diagnostics, which I consider to be quite questionable.* [Participant 17]

Expectations

Overview

Besides *concerns*, we also found *expectations* to be determinants of GPs' attitudes. This category reflects GPs' thoughts and

beliefs about AI-enabled systems' expected benefits and limitations regarding GP care. Although the expected benefits had a positive connotation in the interview data (concepts regarding *diagnostic quality*, *diagnostic efficiency*, and *legal liability*), the expected limitations depicted a negative perspective (concepts encompassing statements relating to a *lack of human competencies* and *time expenditure*).

Diagnostic Quality

Diagnostic quality represents the expectation that AI-enabled systems can improve the quality of care via more accurate and precise diagnosis. It is GPs' job to provide patients with the best possible care, which is why the expected benefits of AI-enabled systems positively influenced the GPs' attitudes. Especially in rare diseases, which GPs do not regularly treat, the expectation from AI-enabled systems is an improvement of diagnostic quality as AI-enabled systems can work with a larger database than the human brain (participant 18). Thus, AI-enabled systems should act as support or a backup for the physician, in parallel or after a medical diagnosis. GP 12 assumed that AI-enabled systems could assist GPs in the decision-making process and thought that this would positively impact the outcome quality:

> *But for rarer diseases, when it comes to making a diagnosis; for example, a red skin spot that I can't classify at all, then it would be conceivable [...] to reaffirm or reassure oneself [by means of AI].* [Participant 12]

Furthermore, the expectation from an AI-enabled system is that it is more enduring than humans. Unlike a physician, an AI-enabled system does not tire and its diagnostic quality does not suffer from human-like, lower-concentration performance during the course of a day. A participant said the following:

> *If AI is well programmed or if there are no failures in it, then AI is more accurate than a person, who is sometimes tired [and thus] makes bad decisions.* [Participant 2]

Diagnostic Efficiency

Besides the expected diagnostic quality, the participants stated that an AI-enabled system's ability to make rapid diagnoses is a further expected benefit. We refer to this expectation as *diagnostic efficiency*. GP 2 transferred the time advantages of using AI-enabled systems to the area of image recognition and expected AI-enabled systems to be 3 times faster than a physician:

> *While a radiologist might manage 60 diagnostic findings a day, the AI could work day and night and deliver perhaps 180 or 200 findings. And if that happens with similar quality, then [...] you could examine many more patients than a human alone could.* [Participant 2]

On the basis of this benefit of AI-enabled systems, GP 14 expected the use of AI-enabled systems to influence disease progression positively. In addition, GP 1 emphasized the necessity of fast-working AI-enabled systems in the detection of health threats:

> *Now a completely new virus has appeared in China or Japan, and to get ahead of it, you need artificial intelligence which can detect [the virus] much faster.* [Participant 1]

Diagnostic efficiency included the GPs' expectations regarding physician support via AI-enabled systems, reducing the daily workload by preselection (participant 7) and patient prioritization (participant 13). This time-saving effort would give GPs some relief and would allow them to concentrate on more serious cases (participant 7).

Legal Liability

Legal liability included the expectation that AI-enabled systems will give GPs legal backing. All decisions will be documented using AI-enabled systems, allowing the providers to prove the correct decision-making approach in a legal proceeding (participant 12). Furthermore, the participants added the assumption that AI-enabled systems could support the physician's choice of treatment. In this context, GP 13 mentioned the following:

> *[With] AI, you can then understand how [the physician] came to a decision because AI said the risk was 0.001.* [Participant 13]

This was supported by the expectation of built-in legal protection and shifting responsibility from the GP *toward* the AI-enabled system (participant 4).

Lack of Human Competencies

Besides the above-mentioned positive determinants, the following *expectations* depicted the perceived limitations of AI-enabled systems. The expected *lack of human competencies* in AI-enabled systems was mentioned with a high emphasis. It included the GPs' assumption that AI-enabled systems do not have certain human competencies, which are, in fact, crucial for adequate and appropriate treatment in GP care. The respondents agreed that AI-enabled systems will not—some said never—be able to have certain human competencies. In this context, empathy (participant 5), intuition (participant 1), gestures (participant 13), experience (participant 12), and clinical reasoning ability (participant 3) were mentioned. These competencies are important in GP care to collect all relevant information to be able to provide optimal care. Of the 18 participants, 2 (11%) participants said the following:

> *There is something behind almost every illness that makes [diagnosis] even more challenging. And if this is not considered, it will not be possible to help a patient comprehensively. And I think [AI] can probably not do this.* [Participant 5]

> *Experience can hardly be replaced by AI. Experience and intuition. And empathy. This is just how I treat people, to get something out of them. So, this is something that defines a good physician and cannot be replaced by AI. Empathy.* [Participant 1]

Furthermore, describing and verbalizing much of the information collected in GP care (such as mimics or gestures) is not always possible. However, it is an essential data input for the proper operation of any technology (participant 13). Participants expressed that many patients just make an appointment to have

some human interaction, for instance, lonely older patients. GP 15 explained this as follows:

> *My experience every day with patients is that they want to be touched, and they want to look you in the eyes.* [Participant 15]

For them, AI-enabled systems seemed to be unable to fulfill these needs. In the context of human competencies, GP 13 underlined AI-enabled systems' limitations:

> *People are certainly beaten by [AI] in many ways. But not in the emotional one.* [Participant 13]

Time Expenditure

Time expenditure included the expectation that in most cases, GPs would need more time for the *decision-making* process by involving AI-enabled systems, because in routine cases, GPs usually diagnose on their own within seconds. In this context, participant 11 commented the following:

> *[...] in routine cases, [AI] would not be a time saver for me.* [Participant 11]

With AI-enabled systems, additional effort is expected by the participants because they fear that data must be entered in the documentation and fed into the AI-enabled system. GP 2 assumed additional time expenditure owing to a person's need to critically reflect on the results of the AI-enabled system.

Environmental Influences

Besides the 2 main categories, we also identified *environmental influences* that influence GPs' attitudes toward AI-enabled systems. The summarized determinants include influences resulting from an evolving working environment (*changing working conditions*), the perspectives and opinions of key stakeholders (*stakeholder influences*), the available IT hardware and software resources (*IT infrastructure*), and the media environment (*media*).

Changing Working Conditions

Changing working conditions included GPs' perspectives on the challenges caused by demographic change (participant 10), a changing spectrum of diseases (participant 1), and the constant increase in medical knowledge (participant 3). Regarding demographic change, GP 1 stated the following:

> *The lack of physicians comes with giant steps, and what is also urgently needed is telemedicine. And this, of course, needs AI with it.* [Participant 1]

However, demographic change also included the necessity to modernize a practice's equipment with new technologies to be interesting for younger physicians (participant 10). AI-enabled systems were also considered necessary to stay updated about the increasing medical knowledge and provide the patients with the best and latest information about their health care (participant 3). Regarding the changing spectrum of germs and viruses and the resulting need for AI-enabled systems, GP 1 referred to the outbreak of the COVID-19 pandemic.

Stakeholder Influences

Another environmental influence was *stakeholder influences*, which indicated how certain groups of people and organizations influence GPs' opinions. The interviews revealed that patients and institutions are key stakeholders in this context. GP 7 said the following:

> *I think we can be influenced [by the patients' opinions] because, in the end, a medical practice follows the market like a small business. If the patients want [AI technologies] and demand [AI technologies], more and more practices will offer it.* [Participant 7]

However, the GPs also stated that they do not expect patients to disapprove of AI-enabled systems (participant 11). In contrast to the patients' opinions, the GPs agreed that the opinions of institutions such as the German Society of General Medicine (Deutsche Gesellschaft für Allgemeinmedizin und Familienmedizin) or the German General Practitioners Association (Deutscher Hausärzteverband) have key roles in the formation of German GPs' attitudes. GPs place trust in these institutions and regard them as scientific and validated committees of their profession (participant 11). Supported by the fact that physicians wish to receive more recommendations on which technologies they should use in practice, the influence of these institutions' attitudes is evident (participant 7).

Media

The concept of *media* referred to all informative sources in which physicians had heard or read about AI-enabled systems. As most participants had not yet worked with AI-enabled systems, we assume that the media strongly contributes to AI literacy, which describes what GPs believe AI is and can do. A participant said the following:

> *Except for what I have read about it in medical journals, [I hardly come in contact with AI].* [Participant 11]

GP 14 suggested that physicians should be informed about AI technology via regular journal articles.

Information Technology Infrastructure

Another factor that influenced attitudes was the often-inadequate *information technology infrastructure* in physicians' practices. In the event of technical problems, AI-enabled systems cannot be used properly or at all, which can undermine optimal patient care. Physicians are skeptical about AI-enabled systems in this regard and prefer the established ways of performing their routines, as they cannot rely on the overall infrastructure, which needs integration of AI technologies to function properly. In this context, a participant mentioned the following:

> *If my system goes down, my AI is on standby, then sorry, I can't diagnose, my system strikes out. That is why it's nice to be able to write down with a pen on paper what a patient has and has received.* [Participant 16]

Individual Characteristics

Environmental influences are external influences, whereas *individual characteristics* are determinants that describe a physician as a person and include character traits, demographic specifics, and knowledge. Although there are many individual characteristics, we found that age and affinity with technology are particularly relevant to the GPs.

Age

The participants who mentioned *age* disagreed on whether it has a role in determining their attitudes. GP 10, an older physician, said the following:

> *I am convinced it needs much work because there is certainly much resistance, which clearly depends on age.* [Participant 10]

Whereas GP 11, a younger physician, stated the following:

> *I also know young colleagues who are my age, and they also have strong reservations [regarding AI].* [Participant 11]

Thus, we included *age* as a relevant characteristic and leave future research endeavors to challenge its influence on a larger scale.

Affinity With Technology

Another influencing factor was *affinity with technology*, which indicated whether being open to new technologies supports a positive attitude toward AI-enabled systems. A participant said the following:

> *Well, there are also people in my generation who were already technically inclined [...]. So, I think that's the key to why people [would use AI] or not.* [Participant 18]

Minimum Requirements of AI-Enabled Systems

Overview

Besides the above-mentioned categories and concepts, the interviews also revealed *the minimum requirements of AI-enabled systems*, which are preconditions that must be met for GPs to contemplate using AI-enabled systems. Although many of the requirements were thematically related to *expectations* and *concerns*, our qualitative data collection allowed us to distinguish between the attitude determinants and the essential and must-have criteria. We will now explain the 6 identified minimum requirements and underline their intensities with statements from the interviews.

Time Efficiency

Most participant statements that expressed demands of AI-enabled systems contributed to the minimum requirement of *time efficiency.* GPs need AI-enabled systems to be fast and easy to use, as they have limited time for each patient consultation. A participant mentioned the following:

> *First of all, [AI] should be fast. There is always time pressure.* [Participant 14]

Also, participants stated that AI-enabled systems must not take additional time, as this would keep a physician from performing essential tasks (participant 15). Thus, the focus was also on practical relevance and system compatibility with existing practice ISs. The participants demanded a self-explanatory design that can be operated quickly and in a few simple steps. The time component's importance in the use of AI-enabled systems was shown by GP 15, who had already tested an AI-enabled system and decided against further use stated as follows:

> *[...] [the use of AI] took me far too long* [Participant 15]

Diagnostic Quality

Besides the time components, *diagnostic quality* was mentioned as another key requirement of AI-enabled systems. For physicians to consider the use of AI-enabled systems, the AI-enabled system must be validated, must not make mistakes, and must provide accurate diagnoses so that there is no threat to patient care (participant 7). Furthermore, some participants demanded accurate diagnoses and even better results through AI-enabled systems compared with human engagement, because otherwise, AI-enabled systems would be obsolete (participant 2). In addition, AI-enabled systems must be evidence-based and must follow guidelines. In this context, GP 10 said the following:

> *[AI must be] scientifically grounded and must provide validated results that [the physician] may not be able to produce in their entirety.* [Participant 10]

Data Security

Participants also named guaranteed *data security* as a requirement for using AI-enabled systems. The physicians justified this requirement with concerns about privacy and misuse of data and they do not want patient and physician data to be accessible to anyone. A participant explained as follows:

> *Of course, it is also important to me that there is corresponding data security. I do not want the patients and us to be completely transparent. That is certainly not in the overall interest.* [Participant 10]

Data security issues were the second reason along with time expenditure that made GP 15 decide to refrain from further using that AI-enabled system.

Economic Viability

Economic viability summarized the statements regarding AI-enabled systems' affordability and questions about financing them. In this regard, GP 2 mentioned the following:

> *If they are affordable [then I would use AI applications].* [Participant 2]

Furthermore, the participants expressed their willingness to use AI-enabled systems based on how the technology is financed and stated that the cost–benefit ratio must be consistent.

Transparency

Transparency and thus the comprehensibility of AI algorithms is another key requirement of AI-enabled systems. To trust AI-enabled systems, it was important to the GPs that the proposals submitted by the AI-enabled system are comprehensible. Thus, a participant said the following:

> *I must know how [AI] obtains information and how [it] works.* [Participant 11]

Autonomy

Autonomy represented another requirement, indicating that an AI-enabled system must be self-managed by the providers. Using the technology is feasible only if a physician can continue to work autonomously and the next treatment steps are not

decided by an AI-enabled system. However, the participants had a negative attitude toward intervention in a physician's self-determined work. A participant explained as follows:

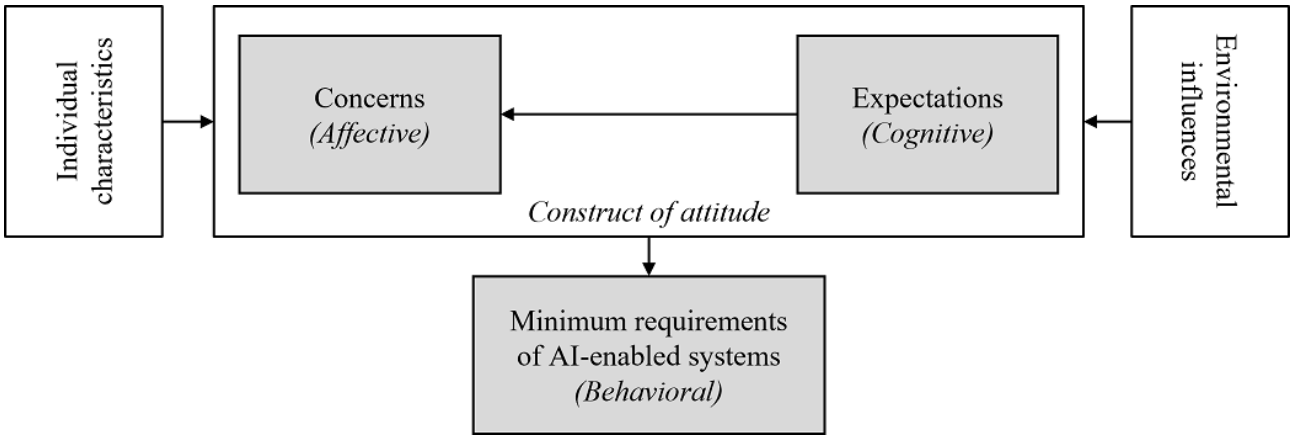
> *I would participate only [on a] voluntarily [basis].* [Participant 15]

Discussion

Principal Findings

We now discuss GPs' *attitude determinants* regarding AI-enabled systems in GP care and the relationships among these determinants. We conflate our findings to propose a model (Figure 2) and derive theoretical and practical contributions. Considering the lack of existing solutions and experiences of GPs with AI-enabled systems, our findings emphasize the relevance of GPs' *AI literacy*. Hence, the interview statements and the resulting discussion are based on the GPs' knowledge of AI, whether this is factual. The results underline that the participating physicians formed an opinion, even if they, as potential end users, did not have the necessary knowledge to understand the technology comprehensively or differentiate AI-enabled systems from knowledge-based CDSSs. Given that this will be the case for a large proportion of solely medically educated GPs, it is more important to investigate the determinants of attitude in rich detail. In doing so, research and practice can derive levers for the successful adoption of AI-enabled systems. Thus, and as the verisimilitude of GPs' AI literacy is debatable, it emphasizes and gives important clues to understanding their attitudes and implications for practice.

Figure 2. Model of the general practitioners' determinants of attitudes toward artificial intelligence (AI)–enabled systems.



Our attitude determinants *concerns*, *expectations*, and *minimum requirements of AI-enabled systems* corroborate the 3-component model of attitude proposed by Rosenberg et al [41]. The data analysis revealed the following: (1) the identified *concerns*, which represent the participants' expressed emotions toward AI-enabled systems and refer to the affective component of attitude; (2) the identified *expectations*, which picture GPs' beliefs toward AI-enabled systems and address the cognitive component; and (3) the identified *minimum requirements of AI-enabled systems*, which are preconditions that must be met for GPs to contemplate using AI-enabled systems and address the behavioral component of attitude.

However, as the relationships among these 3 determinants lack consistency and dependency according to our findings, we could not confirm the 3-component model [43]. Instead, the interviews revealed that GPs' *concerns* and *expectations* form the *minimum requirements of AI-enabled systems*. This approach is consistent with the 2-component model of attitude, which indicates that the affective and cognitive components explain the behavioral intention [44]. For instance, the participants clarified that concerns about data misuse trigger the GPs' demand for data security in AI-enabled systems. The importance of data security in health information technologies is not a novelty but rather a recurring theme in practice and research [71,72]. Another example is the expected time expenditure when using AI-enabled systems, which leads to the requirement that AI-enabled systems must be time efficient and simple to use. As for most health complaints, GP care is the first point of contact and GPs must treat a large number of patients. For instance, in Germany, GP care is one of the most frequently used health care services, with >200 consultations per week per physician [73] and an average physician–patient contact time of 7.6 minutes [26]. Thus, GPs are always under time pressure, which is why every additional action or additional use of new technologies must be well considered [74]. In part, these constraints in GP care are owing to an aging population [74]. GP consultations increase as age correlates with physician visits, particularly in primary care, where a high service use level by the older population is significant [75]. Besides older patients, the aging population also causes an increasing GP shortage owing to retirements and insufficient numbers of successors to GP care [76]. These interdependent developments further reduce the time available for a GP to make an initial diagnosis, which decides whether a patient receives the correct follow-up treatment, is treated at the right time, or receives treatment at all. Thus, a GP's decision strongly impacts the course of treatment and outcome quality [28]. Consequently, increased workload and diagnostic suggestions with the potential to harm patients resulting from the use of AI-enabled systems would likely hamper technology adoption by GPs [77]. On the basis of our findings, we assume that GPs would not use AI-enabled systems if these require additional time or harm patients, despite their benefits. Thus, we consider *diagnostic quality* and *time efficiency* to be the most important *minimum requirements of AI-enabled systems*. Obligations to use AI-enabled systems by regulations or by superiors are neglected in this assumption.

However, we found not only minimum requirements of AI-enabled systems to be influenced by concerns and expectations but also *concerns* and *expectations* to be interrelated and form a construct of attitude. For instance, the participants' concern of being replaceable was caused by their perception of AI-enabled systems formulating more accurate diagnoses than physicians. However, most of our participants did not fear being replaceable, as AI-enabled systems are unable to have and perform human competencies such as empathy and clinical reasoning. Similar conclusions were made by Oh et al [78] who conducted a web-based survey with physicians with the result that most of the participants do not believe that AI will replace physicians. In GP care, decisions are often made with incomplete and fragmented patient-specific information, requiring human competencies such as experience, intuition, and clinical reasoning [79]. Furthermore, in GP care, human competencies are of particular importance to develop a physician–patient relationship. To gather relevant information for the decision-making process, GP care places great importance on interpersonal continuity in the physician–patient relationship [39]. Especially regarding GPs' gatekeeping role and their focus on an emotional bond in medical service provision, this interpersonal relationship is valuable because it enables the therapeutic benefit of improved continuity of care and more holistic and individualized treatments [32]. In summary, interpersonal interaction with patients is very important to GPs, whereas the GPs assume AI-enabled systems to have an insufficient ability to recognize and incorporate important individual aspects gained through the interpersonal relationship. Thus, where and when AI-enabled systems in GP care are useful is to be critically reflected [80]. Considering the potential of AI-enabled systems and their limitations reported in other research streams, we consider hybrid human–AI decision-making a promising scenario to mitigate the weaknesses of each other [36]. Enabling this scenario requires a profound understanding of GPs' barriers to adoption [80], underlining the relevance of our identified attitude determinants.

We also found *individual characteristics* and *environmental influences* to determine GP's attitude toward AI-enabled systems. Regarding individual characteristics, our results for the influence of GP's age are inconclusive. The GPs in our sample presumed that both old and young physicians would have a negative attitude toward AI-enabled systems. However, both old and young participants in our sample generally had a positive attitude toward AI-enabled systems. As this may be owing to a bias in our sampling, we encourage further examinations of age as an attitude determining *individual characteristics*. Regarding *environmental influences*, our respondents indicated that a positive attitude from institutions such as the German General Practitioners Association would positively impact their attitudes toward AI-enabled systems. Moreover, GPs' individual context such as office size and facilities (ie, *information technology infrastructure*) might prove themselves in further studies as determinants for GPs' attitude toward AI-enabled systems. By uncovering *individual characteristics* and *environmental influences* as attitude determinants, we found similarities to the factors *social influence* and *age* of the UTAUT. Albeit, in the UTAUT, these determinants influence the intention to use [48,53,54]. However, in contrast to our findings regarding environmental influences, Jeng and Tzeng [81] concluded that social influence does not affect physicians in Taiwan in adopting CDSSs. This divergence may stem from different cultures, differences in medical education and practice, AI characteristics, and GPs' AI literacy, compared with more established CDSSs. We leave it to future research to further explore these relationships regarding environmental influences.

Furthermore, our findings explicate that the consideration of the affective component of attitude is crucial in the medical context despite being neglected often in well-known theories of behavior and acceptance research [54,82,83]. Our interview data show that GPs' concerns about data privacy and patient safety have high importance in the context of patient care and must not be endangered. AI-enabled systems can mitigate cognitive errors resulting from, among others, GPs' fatigue or distraction [23]. Thus, diagnostic accuracy and patient safety increase [84]. However, at the same time, the integration of AI technologies can also lead to biases such as automation bias [85]. By blindly relying on the AI-enabled systems' suggestions, physicians would no longer critically review them, which can reduce accuracy [86] and increase medical errors [87]. Whether AI-enabled systems promote or minimize cognitive biases depends on how they are used [84]. As AI-enabled systems bear certain concerns, such as the fear of being negatively biased by AI-enabled systems' suggestions, the affective component of attitude also plays a key role in the context of AI. Eventually, the affective component is particularly relevant when investigating GPs' attitudes toward AI-enabled systems. Detecting concerns in the early stage can positively determine GPs' attitudes. When GP care comes into widespread contact with AI technologies, this form of attitude can contribute to a positive intention to use, which in turn lays the foundation for successful implementation.

Besides theoretical contributions, we derived valuable implications for practice by reflecting on GPs' attitudes before the use of AI-enabled systems and familiarization with the technology. We suggest making the topic of AI more prominent in politics, health-related associations, and stakeholder institutions of GP care. Via these institutions, knowledge and education on AI-enabled systems can be offered, thus improving GPs' AI literacy. This allows for the mitigation of concerns such as the *change of the physician–patient relationship* and thus, the diminution of restraints is possible. For this purpose, the distribution of evidence-based information via GP-specific journals and the involvement of advocacy groups are highly recommended, as the GPs value their viewpoints. However, it is also important that potential users are not only informed about the potential of AI technology but also about its limitations and shortcomings on the basis of evidence. In this way, physicians can be empowered to use AI-enabled systems in a reflective manner and thus, for example, prevent automation bias.

Moreover, the identified *minimum requirements of AI-enabled systems* are of particular interest concerning the practical implications. First, AI-enabled systems must be programmed and designed to make its use as easy and fast as possible as stated by participants and widely spread in the literature on user-centricity [88]. Second, AI-enabled systems must be

reliable and free of errors to prevent any harm to patients. In addition, AI-enabled systems must ensure data protection and allow the GP to work autonomously. In addition, politics and health insurance companies should consider monetary subventions for AI-based systems because a remarkable result of the review by Ajami and Bagheri-Tadi [89] is the positive influence of financial support on physicians' willingness to use and engage with technologies [89].

Furthermore, AI-enabled systems may foster so-called *black-box-medicine*, as decisions are less transparent to the patient and to the GP. With this lack of transparency, various types of biases may occur, both for the end users and the AI-enabled system. Such biases may result in patient security, data, and privacy concerns [84,90]. Therefore, along with the responsibility of making an AI-augmented diagnosis, there is also the need to create accountable structures for patient-related outcomes. In a recent study by Khullar et al [91], physicians believed that vendors or the employing health care organizations should be held accountable for AI-induced errors, whereas the general public believed that the physicians themselves should be liable. We see suitable liability regulations and their implications for GP's attitude determinants as a promising field for further research.

Furthermore, AI-enabled systems should be developed to diagnose rare cases because GPs assume that they are faster in routine cases than using AI technology. This information can help developers to narrow the application area and to create better-fitting software solutions. This result also indicates that integrating AI technology is not the solution for every problem. Rather, a critical assessment must be made regarding when using an AI-enabled system makes sense and improves decision-making and when this is not the case. Especially when it comes to human competencies and interpersonal relationships, AI-enabled systems cannot replace GPs. Rather, AI-enabled systems should be designed to free up GPs' time so that they have more time to nurture relationships with their patients, which is of particular relevance for diagnosis in GP care. Our findings may serve for a better understanding of how to design AI-enabled systems in a conducive manner and how to foster GP's acceptance in the later adoption of such systems.

Limitations and Future Research

Although we rigorously followed our designed research approach, our study has limitations, some of which are bound to the choice of a qualitative-explorative approach. By design, qualitative interviews do not focus on drawing conclusions for entire populations, which affects the generalizability of the results. Nevertheless, a qualitative approach is appropriate before a quantitative study when dealing with a new and emotionally charged topic. This approach is reinforced by recent research that puts traditional IS adoption models to the test for AI; thus, calling for in-depth reflections [21]. Blease et al [39] also recommended a qualitative approach, as they reported lack of detailed information on GPs' views of AI-enabled systems owing to their quantitative approach. Furthermore, conducting interviews just in 1 country, more precisely, in 1 geographic area within that country might be a limitation of our study. As depicted in existing research, attitudes toward technology might differ between people living in rural areas and urban areas [92]. We recommend collecting data in other countries and conducting cross-country studies to detect differences among these settings. Furthermore, GPs' mostly basic AI literacy is another limitation of our study. Although all study participants were given the same definition of AI-enabled systems at the start of the interview, their statements reflect different understandings. However, the early consideration of the GPs' attitudes, regardless of their technical knowledge is important to identify barriers to implementation at an early stage and derive basic conclusions for AI system design. We must also assume that only GPs who are interested in AI-enabled systems might have a general affinity with technology or who have a strong opinion on AI agreed to be interviewed. This could also explain why none of the participants had a solely negative attitude toward AI-enabled systems.

We further suggest examining the role of the affective attitude component, because we revealed the importance of the identified concerns in our study; whereas, in well-known theories of technology acceptance, this component is often neglected. A closer examination of the affective component will make it possible to determine the extent to which it is relevant in the medical and IS contexts.

Conclusions

AI-enabled systems are considered as promising solutions to enhance both the effectiveness and quality of health care. Especially in GP care, which is the first point of contact for most medical needs, physicians deal with a shrinking physician–patient time and incomplete or sometimes incorrect information. Here, AI technology promises new solutions to support physicians and decrease diagnostic errors that lead to extensive consequences. Although the application potential of AI-enabled systems in health care has been widely discussed theoretically and conceptually, a widespread application in the professional practice of GPs is still dreams of the future. To tap the undisputed potential of AI-enabled systems in practical use, a fundamental investigation of the technical systems and social actors is required. As academic research, in this respect, is still in its infancy, we investigated the attitudes of GPs toward AI-enabled systems. Thereby, we seek to contribute to a better understanding of GPs' attitudes, which is crucial for developing and implementing suitable AI-enabled systems. Thus, we used in-depth qualitative-explorative interview data with German GPs and proposed a preliminary research model. We identified three determinants of GPs' attitudes: *concerns*, *expectations*, and *minimum requirements of AI-enabled systems*. Furthermore, we revealed *individual characteristics* and *environmental influences* as the 2 conditional determinants of GPs' attitudes toward AI-enabled systems. The findings emphasize the importance of attitude's affective component at the interface of medical and AI research. Moreover, the findings show that diagnostic quality and time efficiency are mandatory for GPs to even consider the use of AI-enabled systems. Therefore, integrating user groups' attitudes and needs is a fundamental prerequisite for user-centered design, which leads to a higher willingness and inclusion of the systems into everyday use. Considering that the GPs in our interview study predominantly corroborated AI-enabled systems' seminal role in the future of

GP care, our findings may serve as a foundation for future research. Besides investigating the attitudes of user groups in other fields in the health care system, research endeavors should also focus on how the attitudes of GPs toward AI-enabled systems can be proactively promoted. In addition, future work should include and conflate findings from related research areas such as human-computer interaction, psychology, sociology, and computer science to account for AI's interdisciplinary implications for health care.

Acknowledgments

This study was funded by the University of Bayreuth Open Access Publishing Fund.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Interview guideline.
[DOCX File , 18 KB - jmir_v24i1e28916_app1.docx]

References

1. Frank MR, Autor D, Bessen JE, Brynjolfsson E, Cebrian M, Deming DJ, et al. Toward understanding the impact of artificial intelligence on labor. Proc Natl Acad Sci U S A 2019 Apr 02;116(14):6531-6539 [FREE Full text] [doi: 10.1073/pnas.1900949116] [Medline: 30910965]
2. Ågerfalk PJ. Artificial intelligence as digital agency. Eur J Inf Syst 2020 Feb 17;29(1):1-8. [doi: 10.1080/0960085x.2020.1721947]
3. Bundy A. Preparing for the future of artificial intelligence. AI Soc 2016 Nov 29;32(2):285-287. [doi: 10.1007/s00146-016-0685-0]
4. Jiang F, Jiang Y, Zhi H, Dong Y, Li H, Ma S, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol 2017 Dec;2(4):230-243 [FREE Full text] [doi: 10.1136/svn-2017-000101] [Medline: 29507784]
5. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science 2015 Jul 17;349(6245):255-260. [doi: 10.1126/science.aaa8415] [Medline: 26185243]
6. Russel S, Norvig P. Artificial Intelligence: A Modern Approach, 3rd edition. Boston: Pearson; 2016.
7. Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery. Lancet Oncol 2019 May;20(5):e262-e273. [doi: 10.1016/S1470-2045(19)30149-4] [Medline: 31044724]
8. Pham BT, Nguyen MD, Bui KT, Prakash I, Chapi K, Bui DT. A novel artificial intelligence approach based on Multi-layer Perceptron Neural Network and Biogeography-based Optimization for predicting coefficient of consolidation of soil. CATENA 2019 Feb;173:302-311. [doi: 10.1016/j.catena.2018.10.004]
9. Rai A, Constantinides P, Sarker S. Next generation digital platforms : toward human-AI hybrids. Digital Infrastructures & Platforms. 2019. URL: https://tinyurl.com/v6vc8kph [accessed 2022-01-13]
10. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Future Healthc J 2019 Jun;6(2):94-98 [FREE Full text] [doi: 10.7861/futurehosp.6-2-94] [Medline: 31363513]
11. Fowler A. The role of AI-based technology in support of the knowledge management value activity cycle. J Strategic Inf Syst 2000 Sep;9(2-3):107-128. [doi: 10.1016/s0963-8687(00)00041-x]
12. Castaneda C, Nalley K, Mannion C, Bhattacharyya P, Blake P, Pecora A, et al. Clinical decision support systems for improving diagnostic accuracy and achieving precision medicine. J Clin Bioinforma 2015;5:4 [FREE Full text] [doi: 10.1186/s13336-015-0019-3] [Medline: 25834725]
13. Hersh WR. Medical informatics: improving health care through information. JAMA 2002;288(16):1955-1958. [doi: 10.1001/jama.288.16.1955] [Medline: 12387634]
14. Bright TJ, Wong A, Dhurjati R, Bristow E, Bastian L, Coeytaux RR, et al. Effect of clinical decision-support systems: a systematic review. Ann Intern Med 2012 Jul 03;157(1):29-43 [FREE Full text] [doi: 10.7326/0003-4819-157-1-201207030-00450] [Medline: 22751758]
15. Jaspers MW, Smeulers M, Vermeulen H, Peute LW. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings. J Am Med Inform Assoc 2011 May 01;18(3):327-334 [FREE Full text] [doi: 10.1136/amiajnl-2011-000094] [Medline: 21422100]
16. Khairat S, Marc D, Crosby W, Al Sanousi A. Reasons for physicians not adopting clinical decision support systems: critical analysis. JMIR Med Inform 2018 Apr 18;6(2):e24 [FREE Full text] [doi: 10.2196/medinform.8912] [Medline: 29669706]
17. Sim I, Gorman P, Greenes RA, Haynes RB, Kaplan B, Lehmann H, et al. Clinical decision support systems for the practice of evidence-based medicine. J Am Med Inform Assoc 2001;8(6):527-534 [FREE Full text] [doi: 10.1136/jamia.2001.0080527] [Medline: 11687560]

18. Sutton RT, Pincock D, Baumgart DC, Sadowski DC, Fedorak RN, Kroeker KI. An overview of clinical decision support systems: benefits, risks, and strategies for success. NPJ Digit Med 2020;3:17 [FREE Full text] [doi: 10.1038/s41746-020-0221-y] [Medline: 32047862]
19. Legg S, Hutter M. Universal intelligence: a definition of machine intelligence. Minds Mach 2007 Nov 10;17(4):391-444. [doi: 10.1007/s11023-007-9079-x]
20. Russel S. Human Compatible: Artificial Intelligence and the Problem of Control. London, United Kingdom: Penguin Books; 2019.
21. Schuetz S, Venkatesh V. "Research perspectives: the rise of human machines: how cognitive computing systems challenge assumptions of user-system interaction ". J Assoc Inf Syst 2020;21(1):460-482. [doi: 10.17705/1jais.00608]
22. The Lancet. Artificial intelligence in health care: within touching distance. Lancet 2017 Dec 23;390(10114):2739. [doi: 10.1016/S0140-6736(17)31540-4] [Medline: 29303711]
23. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019 Jan;25(1):44-56. [doi: 10.1038/s41591-018-0300-7] [Medline: 30617339]
24. Yu K, Beam AL, Kohane IS. Artificial intelligence in healthcare. Nat Biomed Eng 2018 Oct;2(10):719-731. [doi: 10.1038/s41551-018-0305-z] [Medline: 31015651]
25. Pannu A. Artificial intelligence and its application in different areas. Int J Eng Innov Technol 2015 Apr 10;4(10):79-84 [FREE Full text]
26. Deveugele M, Derese A, van den Brink-Muinen A, Bensing J, De Maeseneer J. Consultation length in general practice: cross sectional study in six European countries. BMJ 2002 Aug 31;325(7362):472 [FREE Full text] [doi: 10.1136/bmj.325.7362.472] [Medline: 12202329]
27. Singh H, Giardina TD, Meyer AN, Forjuoh SN, Reis MD, Thomas EJ. Types and origins of diagnostic errors in primary care settings. JAMA Intern Med 2013 Mar 25;173(6):418-425 [FREE Full text] [doi: 10.1001/jamainternmed.2013.2777] [Medline: 23440149]
28. Van Such M, Lohr R, Beckman T, Naessens JM. Extent of diagnostic agreement among medical referrals. J Eval Clin Pract 2017 Aug;23(4):870-874. [doi: 10.1111/jep.12747] [Medline: 28374457]
29. Lambe KA, O'Reilly G, Kelly B, Curristan S. Dual-process cognitive interventions to enhance diagnostic reasoning: a systematic review. BMJ Qual Saf 2016 Oct;25(10):808-820. [doi: 10.1136/bmjqs-2015-004417] [Medline: 26873253]
30. Police R, Foster T, Wong K. Adoption and use of health information technology in physician practice organisations: systematic review. Inform Prim Care 2010;18(4):245-258 [FREE Full text] [doi: 10.14236/jhi.v18i4.780] [Medline: 22040851]
31. Aronson SJ, Rehm HL. Building the foundation for genomics in precision medicine. Nature 2015 Oct 15;526(7573):336-342 [FREE Full text] [doi: 10.1038/nature15816] [Medline: 26469044]
32. Harbishettar V, Krishna K, Srinivasa P, Gowda M. The enigma of doctor-patient relationship. Indian J Psychiatry 2019 Apr;61(Suppl 4):S776-S781 [FREE Full text] [doi: 10.4103/psychiatry.IndianJPsychiatry_96_19] [Medline: 31040473]
33. Nurek M, Kostopoulou O, Delaney BC, Esmail A. Reducing diagnostic errors in primary care. A systematic meta-review of computerized diagnostic decision support systems by the LINNEAUS collaboration on patient safety in primary care. Eur J Gen Pract 2015 Sep;21 Suppl:8-13 [FREE Full text] [doi: 10.3109/13814788.2015.1043123] [Medline: 26339829]
34. Bryan C, Boren S. The use and effectiveness of electronic clinical decision support tools in the ambulatory/primary care setting: a systematic review of the literature. Inform Prim Care 2008;16(2):79-91 [FREE Full text] [doi: 10.14236/jhi.v16i2.679] [Medline: 18713524]
35. Asan O, Bayrak AE, Choudhury A. Artificial intelligence and human trust in healthcare: focus on clinicians. J Med Internet Res 2020 Jun 19;22(6):e15154 [FREE Full text] [doi: 10.2196/15154] [Medline: 32558657]
36. Dellermann D, Ebel P, Söllner M, Leimeister JM. Hybrid intelligence. Bus Inf Syst Eng 2019 Mar 28;61(5):637-643. [doi: 10.1007/s12599-019-00595-2]
37. Bhattacherjee A, Hikmet N. Physicians' resistance toward healthcare information technology: a theoretical model and empirical test. Eur J Inf Syst 2017 Dec 19;16(6):725-737. [doi: 10.1057/palgrave.ejis.3000717]
38. Liberati EG, Ruggiero F, Galuppo L, Gorli M, González-Lorenzo M, Maraldi M, et al. What hinders the uptake of computerized decision support systems in hospitals? A qualitative study and framework for implementation. Implement Sci 2017 Sep 15;12(1):113 [FREE Full text] [doi: 10.1186/s13012-017-0644-2] [Medline: 28915822]
39. Blease C, Kaptchuk TJ, Bernstein MH, Mandl KD, Halamka JD, DesRoches CM. Artificial intelligence and the future of primary care: exploratory qualitative study of UK general practitioners' views. J Med Internet Res 2019 Mar 20;21(3):e12802 [FREE Full text] [doi: 10.2196/12802] [Medline: 30892270]
40. Eagly A. Psychology of Attitudes. Belmont, CA: Wadsworth Publishing; 1993.
41. Rosenberg MJ, Hovland CI, McGuire WJ, Abelson RP, Brehm JW. Attitude organization and change: An analysis of consistency among attitude components. Westport, Connecticut, United States: Praeger; 1960.
42. Bierhoff HW. Sozialpsychologie: Ein Lehrbuch, 6th edition. Stuttgart: Kohlhammer; 2006.
43. Fischer L, Wiswede G. Grundlagen der Sozialpsychologie, 3rd edition. Berlin: Walter de Gruyter; 2014.
44. Bagozzi RP, Burnkrant RE. Attitude measurement and behavior change: a reconsideration of attitude organization and its relationship to behavior. Advances Consum Res 1979;6(1):295-302.

45. Bagozzi RP, Burnkrant RE. Attitude organization and the attitude–behavior relationship. J Personality Soc Psychol 1979 Jun;37(6):913-929. [doi: 10.1037/0022-3514.37.6.913]
46. Bagozzi RP. Attitudes, intentions, and behavior: a test of some key hypotheses. J Personality Soc Psychol 1981;41(4):607-627. [doi: 10.1037/0022-3514.41.4.607]
47. Fischbein M, Ajzen I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading: Addison-Wesley; 1975.
48. Ajzen I. The theory of planned behavior. Organizational Behav Human Decision Proc 1991 Dec;50(2):179-211. [doi: 10.1016/0749-5978(91)90020-T]
49. Ajzen I, Fishbein M. The influence of attitudes on behavior. In: The Handbook of Attitudes. East Sussex, United Kingdom: Psychology Press; 2005.
50. Hussein Z, Oon SW, Fikry A. Consumer attitude: does it influencing the intention to use mHealth? Procedia Comput Sci 2017;105:340-344. [doi: 10.1016/j.procs.2017.01.231]
51. Pinto Dos Santos D, Giese D, Brodehl S, Chon SH, Staab W, Kleinert R, et al. Medical students' attitude towards artificial intelligence: a multicentre survey. Eur Radiol 2019 Apr;29(4):1640-1646. [doi: 10.1007/s00330-018-5601-1] [Medline: 29980928]
52. Sit C, Srinivasan R, Amlani A, Muthuswamy K, Azam A, Monzon L, et al. Attitudes and perceptions of UK medical students towards artificial intelligence and radiology: a multicentre survey. Insights Imaging 2020 Feb 05;11(1):14 [FREE Full text] [doi: 10.1186/s13244-019-0830-7] [Medline: 32025951]
53. Ajzen I, Fishbein M. Understanding Attitudes and Predicting Social Behavior. London, United Kingdom: Pearson; 1980.
54. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: toward a unified view. MIS Q 2003;27(3):425-478. [doi: 10.2307/30036540]
55. Venkatesh V, Thong J, Xu X. Unified theory of acceptance and use of technology: a synthesis and the road ahead. J Assoc Inf Syst 2016 May;17(5):328-376. [doi: 10.17705/1jais.00428]
56. Hirschheim R. Introduction to the special issue on “Quo Vadis TAM – Issues and reflections on technology acceptance research”. JAIS 2007 Apr;8(4):203-205. [doi: 10.17705/1jais.00128]
57. Walsham G. Interpreting Information Systems in Organizations. Chichester: John Wiley & Sons; Jan 1993.
58. Kaplan B, Maxwell JA. Qualitative research methods for evaluating computer information systems. In: Evaluating the Organizational Impact of Healthcare Information Systems. New York: Springer; 2005.
59. Polkinghorne DE. Language and meaning: data collection in qualitative research. J Counsel Psychol 2005 Apr;52(2):137-145. [doi: 10.1037/0022-0167.52.2.137]
60. Schultze U, Avital M. Designing interviews to generate rich data for information systems research. Inf Organization 2011 Jan;21(1):1-16. [doi: 10.1016/j.infoandorg.2010.11.001]
61. Nelson J. Using conceptual depth criteria: addressing the challenge of reaching saturation in qualitative research. Qual Res 2016 Dec 14;17(5):554-570. [doi: 10.1177/1468794116679873]
62. Myers MD, Newman M. The qualitative interview in IS research: examining the craft. Inf Organization 2007 Jan;17(1):2-26. [doi: 10.1016/j.infoandorg.2006.11.001]
63. Castillo-Montoya M. Preparing for interview research: the interview protocol refinement framework. Qual Report 2016 May 1;21(5):811-831. [doi: 10.46743/2160-3715/2016.2337]
64. Ryan F, Coughlan M, Cronin P. Interviewing in qualitative research: the one-to-one interview. Int J Ther Rehab 2009 Jun;16(6):309-314. [doi: 10.12968/ijtr.2009.16.6.42433]
65. Glaser BG. Theoretical Sensitivity: Advances in the Methodology of Grounded Theory. Mill Valley: Sociology Press; 1978.
66. Corbin JM, Strauss A. Basics of Qualitative Research Techniques and Procedures for Developing Grounded Theory. Thousand Oaks, California, United States: Sage Publications; 2014.
67. Kallio H, Pietilä AM, Johnson M, Kangasniemi M. Systematic methodological review: developing a framework for a qualitative semi-structured interview guide. J Adv Nurs 2016 Dec;72(12):2954-2965. [doi: 10.1111/jan.13031] [Medline: 27221824]
68. Glaser BG, Strauss A. Discovery of Grounded Theory Strategies for Qualitative Research. Oxfordshire, England, UK: Routledge; 2000.
69. Etikan I. Comparison of convenience sampling and purposive sampling. Am J Theoretical Appl Stat 2016 Jun;5(1):1-11. [doi: 10.11648/j.ajtas.20160501.11]
70. Long D, Magerko B. What is AI literacy? Competencies and design considerations. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. 2020 Presented at: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems; Apr 25 - 30, 2020; Honolulu HI USA. [doi: 10.1145/3313831.3376727]
71. de Grood C, Raissi A, Kwon Y, Santana M. Adoption of e-health technology by physicians: a scoping review. J Multidiscip Healthc 2016;9:335-344 [FREE Full text] [doi: 10.2147/JMDH.S103881] [Medline: 27536128]
72. Steininger K, Stiglbauer B. EHR acceptance among Austrian resident doctors. Health Policy Technol 2015 Jun;4(2):121-130. [doi: 10.1016/j.hlpt.2015.02.003]

73. Irving G, Neves AL, Dambha-Miller H, Oishi A, Tagashira H, Verho A, et al. International variations in primary care physician consultation time: a systematic review of 67 countries. BMJ Open 2017 Nov 08;7(10):e017902 [FREE Full text] [doi: 10.1136/bmjopen-2017-017902] [Medline: 29118053]
74. Reed M, Lehmann B, Herrmann M. The evolving state of general practice and GP education in Germany. Health Care Current Rev 2017;05(03):1-4. [doi: 10.4172/2375-4273.1000203]
75. Nie JX, Wang L, Tracy C, Moineddin R, Upshur RE. Health care service utilization among the elderly: findings from the Study to Understand the Chronic Condition Experience of the Elderly and the Disabled (SUCCEED project). J Eval Clin Pract 2008 Dec;14(6):1044-1049. [doi: 10.1111/j.1365-2753.2008.00952.x] [Medline: 19019098]
76. Natanzon I, Ose D, Szecsenyi J, Campbell S, Roos M, Joos S. Does GPs' self-perception of their professional role correspond to their social self-image?--a qualitative study from Germany. BMC Fam Pract 2010 Feb 04;11:10 [FREE Full text] [doi: 10.1186/1471-2296-11-10] [Medline: 20132534]
77. Young AJ. New technologies and general practice. Br J Gen Pract 2016 Nov 24;66(653):601-602. [doi: 10.3399/bjgp16x688021]
78. Oh S, Kim JH, Choi S, Lee HJ, Hong J, Kwon SH. Physician confidence in artificial intelligence: an online mobile survey. J Med Internet Res 2019 Mar 25;21(3):e12422 [FREE Full text] [doi: 10.2196/12422] [Medline: 30907742]
79. Yazdani S, Hosseinzadeh M, Hosseini F. Models of clinical reasoning with a focus on general practice: a critical review. J Adv Med Educ Prof 2017 Oct;5(4):177-184 [FREE Full text] [Medline: 28979912]
80. Knop M, Mueller M, Freude H. Perceived limitations of telemedicine from a phenomenological perspective. In: Proceedings of the 33rd Bled eConference. 2020 Presented at: Proceedings of the 33rd Bled eConference; Jun 28 - Jul 1 2020; Bled, Slovenia. [doi: 10.18690/978-961-286-362-3.9]
81. Jeng DJ, Tzeng G. Social influence on the use of clinical decision support systems: revisiting the unified theory of acceptance and use of technology by the fuzzy Dematel technique. Comput Industrial Eng 2012 Apr;62(3):819-828. [doi: 10.1016/j.cie.2011.12.016]
82. Davis FD, Bagozzi RP, Warshaw PR. Extrinsic and intrinsic motivation to use computers in the workplace1. J Appl Social Pyschol 1992 Jul;22(14):1111-1132. [doi: 10.1111/j.1559-1816.1992.tb00945.x]
83. Kulviwat S, Bruner II GC, Kumar A, Nasco SA, Clark T. Toward a unified theory of consumer acceptance technology. Psychol Mark 2007 Dec;24(12):1059-1084. [doi: 10.1002/mar.20196]
84. Felmingham CM, Adler NR, Ge Z, Morton RL, Janda M, Mar VJ. The importance of incorporating human factors in the design and implementation of artificial intelligence for skin cancer diagnosis in the real world. Am J Clin Dermatol 2021 Mar;22(2):233-242. [doi: 10.1007/s40257-020-00574-4] [Medline: 33354741]
85. Bond RR, Novotny T, Andrsova I, Koc L, Sisakova M, Finlay D, et al. Automation bias in medicine: the influence of automated diagnoses on interpreter accuracy and uncertainty when reading electrocardiograms. J Electrocardiol 2018;51(6S):S6-11. [doi: 10.1016/j.jelectrocard.2018.08.007] [Medline: 30122457]
86. Tsai T, Fridsma D, Gatti G. Computer decision support as a source of interpretation error: the case of electrocardiograms. J Am Med Inform Assoc 2003;10(5):478-483 [FREE Full text] [doi: 10.1197/jamia.M1279] [Medline: 12807810]
87. Lyell D, Magrabi F, Raban MZ, Pont L, Baysari MT, Day RO, et al. Automation bias in electronic prescribing. BMC Med Inform Decis Mak 2017 Mar 16;17(1):28 [FREE Full text] [doi: 10.1186/s12911-017-0425-5] [Medline: 28302112]
88. Nguyen BV, Burstein F, Fisher J. Improving service of online health information provision: a case of usage-driven design for health information portals. Inf Syst Front 2014 May 27;17(3):493-511. [doi: 10.1007/s10796-014-9507-4]
89. Ajami S, Bagheri-Tadi T. Barriers for adopting Electronic Health Records (EHRs) by physicians. Acta Inform Med 2013;21(2):129-134 [FREE Full text] [doi: 10.5455/aim.2013.21.129-134] [Medline: 24058254]
90. Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc 2020 Mar 01;27(3):491-497 [FREE Full text] [doi: 10.1093/jamia/ocz192] [Medline: 31682262]
91. Khullar D, Casalino L, Qian Y, Lu Y, Chang E, Aneja S. Public vs physician views of liability for artificial intelligence in health care. J Am Med Inform Assoc 2021 Jul 14;28(7):1574-1577. [doi: 10.1093/jamia/ocab055] [Medline: 33871009]
92. Müller M, Knop M, Ressing C. Constituting factors of a digitally influenced relationship between patients and primary care physicians in rural areas. In: Proceedings of the Hawaii International Conference on System Sciences. 2020 Presented at: Hawaii International Conference on System Sciences; Jan 7-10, 2020; MAUI, HI, USA. [doi: 10.24251/hicss.2020.447]

Abbreviations

AI: artificial intelligence
CDSS: clinical decision support system
DSS: decision support system
GP: general practitioner
GTM: grounded theory methodology
IS: information system
UTAUT: unified theory of acceptance and use of technology

Edited by R Kukafka; submitted 18.03.21; peer-reviewed by D Vogel, M Knop; comments to author 31.03.21; revised version received 24.06.21; accepted 21.11.21; published 27.01.22.

Please cite as:
Buck C, Doctor E, Hennrich J, Jöhnk J, Eymann T
General Practitioners' Attitudes Toward Artificial Intelligence–Enabled Systems: Interview Study
J Med Internet Res 2022;24(1):e28916
URL: https://www.jmir.org/2022/1/e28916
doi:10.2196/28916
PMID:35084342

Original Paper

An Artificial Intelligence Chatbot for Young People's Sexual and Reproductive Health in India (SnehAI): Instrumental Case Study

Hua Wang[1], PhD; Sneha Gupta[1], MBA; Arvind Singhal[2,3], PhD; Poonam Muttreja[4], MPA; Sanghamitra Singh[4], PhD; Poorva Sharma[5], BSc; Alice Piterova[5], MA

[1]Department of Communication, University at Buffalo, The State University of New York, Buffalo, NY, United States

[2]Department of Communication, The University of Texas at El Paso, El Paso, TX, United States

[3]School of Business and Social Sciences, Inland University of Applied Sciences, Elverum, Norway

[4]Population Foundation of India, New Delhi, India

[5]AI for Good UK, London, United Kingdom

Corresponding Author:
Hua Wang, PhD
Department of Communication
University at Buffalo, The State University of New York
359 Baldy Hall
Buffalo, NY, 14260
United States
Phone: 1 7166451501
Email: hwang23@buffalo.edu

Abstract

Background: Leveraging artificial intelligence (AI)–driven apps for health education and promotion can help in the accomplishment of several United Nations sustainable development goals. SnehAI, developed by the Population Foundation of India, is the first Hinglish (Hindi + English) AI chatbot, deliberately designed for social and behavioral changes in India. It provides a private, nonjudgmental, and safe space to spur conversations about taboo topics (such as safe sex and family planning) and offers accurate, relatable, and trustworthy information and resources.

Objective: This study aims to use the Gibson theory of affordances to examine SnehAI and offer scholarly guidance on how AI chatbots can be used to educate adolescents and young adults, promote sexual and reproductive health, and advocate for the health entitlements of women and girls in India.

Methods: We adopted an instrumental case study approach that allowed us to explore SnehAI from the perspectives of technology design, program implementation, and user engagement. We also used a mix of qualitative insights and quantitative analytics data to triangulate our findings.

Results: SnehAI demonstrated strong evidence across fifteen functional affordances: accessibility, multimodality, nonlinearity, compellability, queriosity, editability, visibility, interactivity, customizability, trackability, scalability, glocalizability, inclusivity, connectivity, and actionability. SnehAI also effectively engaged its users, especially young men, with 8.2 million messages exchanged across a 5-month period. Almost half of the incoming user messages were texts of deeply personal questions and concerns about sexual and reproductive health, as well as allied topics. Overall, SnehAI successfully presented itself as a trusted friend and mentor; the curated content was both entertaining and educational, and the natural language processing system worked effectively to personalize the chatbot response and optimize user experience.

Conclusions: SnehAI represents an innovative, engaging, and educational intervention that enables vulnerable and hard-to-reach population groups to talk and learn about sensitive and important issues. SnehAI is a powerful testimonial of the vital potential that lies in AI technologies for social good.

(J Med Internet Res 2022;24(1):e29969) doi:10.2196/29969

KEYWORDS

artificial intelligence; chatbot; Facebook; affordance; sex education; sexual and reproductive health; contraception; case study; young people; India; transmedia; mobile apps; mobile health; technology design; user engagement; digital health; mobile phone

Introduction

Background

This paper presents rich insights from an instrumental case study of an innovative chatbot in India called SnehAI, which was purposefully conceptualized and designed by the Population Foundation of India to educate and inspire adolescents and young adults to live healthy lives, promote sexual and reproductive health (SRH), and advocate for the health and well-being of women and girls. The SnehAI chatbot aims to provide a safe space for Indian youth to have conversations about SRH, dispel sex-related myths and taboos, offer accurate information about safe sex and contraceptive choices, and address mental health concerns. With a population of approximately 1.4 billion, India accounts for about 18% of all people on the planet [1], with half of this population being under the age of 25 years [2]. Despite stated policy commitments and significant strides made in recent years, the informational needs of adolescents and youth are poorly met, quality education about SRH is highly limited, contraceptive practices are heavily skewed toward female sterilization, and unsafe abortions are rampant [3-6].

In particular, young people in India have limited awareness of contraception and sexually transmitted infections; their knowledge base consists of inaccurate information; and their family life education is highly insufficient [7]. Although the government-endorsed national adolescent health program Rashtriya Kishor Swasthya Karyakram (RKSK) has included SRH as part of its mandate since 2014, direct contact with frontline health workers, even by married young women, was extremely low. Contact with unmarried youth and use of SRH services at adolescent-friendly health clinics are almost completely amiss [8]. Uncomfortable and embarrassed to ask, young people in India have increasingly referred to web-based platforms to look for answers to SRH questions and have garnered misleading or incorrect information [9,10]. In a day and age when mobile services and social media are proliferating in India, this is unfortunate. At the beginning of 2020, India boasted of around 1.1 billion mobile phone connections, covering 78% of the population [11]. With attractive pricing from India-based telecom giants, such as Jio, internet penetration and social media use through mobile networks are rapidly growing [11,12]. Facebook (Meta Platforms) is an obvious leader in the social media space in India, with 320 million users [13]. With this massive expansion of information infrastructure comprising wireless networks, digital technologies, and social media, Indian youths, both in urban and rural areas, are increasingly being plugged into this technology web.

One technology within the realm of social media that has experienced a rapid rise in different industries is chatbots [14]. Chatbots are automated nonhuman agents that engage in conversations with human actors [14]. By design, the user experience in a chatbot strives to be pleasant, as it mimics a scenario in which 2 humans are talking with each other. A chatbot responds by accessing information stored in large digital data repositories. Chatbots quickly sieve what is relevant and convert programing codes into expressions that humans can understand. Although chatbots are often text based, their capabilities have exploded since the pioneering program ELIZA [15], especially in their increased sophistication and accuracy in understanding natural language using artificial intelligence (AI) technologies [16,17]. With AI, chatbots create a dynamic library of answers, building on their existing database with each conversation that occurs. This process of machine learning makes their deployment immensely valuable for disseminating information. Not surprisingly, AI chatbot health apps are burgeoning [18-21], and they have unique advantages for addressing sensitive and taboo issues such as SRH [22]. However, to date, AI chatbot initiatives in low-income countries are scarce. Furthermore, for chatbot projects that do get underway, systematic documentation and assessment are needed to generate new knowledge for the greater public good [23].

Winner of the 2020 eNGO Challenge Award in Digital Tools and Empowerment [24], SnehAI represents the first Hinglish (Hindi + English) AI chatbot that is cocreated *with* and *for* young people, especially those from vulnerable sections of society, to deliberately facilitate communication about SRH topics and promote social and behavioral change [25,26]. It benefits the target population in many unique ways, but most importantly, it fills a gap that exists in the information around SRH. In a world where digital technologies for underdeveloped markets fail to consider consumers' hedonic needs, such as play, romance, and entertainment, SnehAI provides an unusual repository of educational knowledge in an entertaining and engaging container, commonly referred to as *entertainment-education* or *edutainment* [27,28]. Recent anthropological studies have shown that digital natives, irrespective of their socioeconomic status, tend to be attracted to entertainment and storytelling when consuming social media content [23]. Thus, to spread crucial health information in diverse populations, the content of social media platforms needs to engage users. Therefore, the SnehAI chatbot in India offers a terrific opportunity for systematic investigation.

We focused our investigation on the different ways that SnehAI enables SRH information sharing and user engagement. First, we presented our guiding framework, the theory of affordances. Our literature review highlights both foundational arguments and affordances of relevant technologies. We then described our methodological approach, an instrumental case study that harvests the unique contributions of SnehAI. Our findings are organized by our theoretically derived research questions: first, on the functional affordances of SnehAI, and second, on the user engagement patterns of SnehAI. We concluded with a discussion of these major findings, raising implications for theory and practice, and set some directions for future endeavors.

The Gibson Theory of Affordances

The verb to *afford* means to make available, to provide, or to offer. The noun *affordance* did not exist until ecologic psychologist Gibson [29] introduced his theory of affordances. He argued that animals (including humans) and their living environments are one world, and we cannot understand it fully if we treat them as separate entities. Essentially, we are all created by the world we live in, and our behaviors are shaped by our environment [30]. The relational dynamics between us

and our environment can enable or prohibit certain action possibilities [30-32]. Presently, a smartphone can allow us to make calls, send messages, browse the web, and guide us to locations; however, without a charged battery or a correct password, nothing can be done. Gibson [30] also pointed out that affordances may easily imply a positive connotation; however, it is important to remember that they can also be negative. Certain possibilities may be enabling whereas others are constraining [31]. Presently, a mass email can easily reach several people with a click of the *send* button, but if infected, it can also spread a computer virus across the world in no time. Moreover, Gibson [30] emphasized that the affordances of an environment exist permanently—some may be latent that require discovery, but most should be directly perceivable without extra effort of learning. Affordances invite concrete behaviors, which, in turn, can afford other behaviors [30] and, in some cases, provide agency to seek value and meaning [33,34]. Most young people can figure out how to navigate various social networking sites rather intuitively; and given that their behaviors are observable to others, this can trigger additional web-based behaviors.

Functional Affordances of Technologies

Over the past 4 decades, the theory of affordances has deeply influenced the design of everyday objects and user experiences of human–computer interaction [35-37]. Affordances have been used as a high-level theoretical framework to understand the internet [38,39], mobile and social media [40,41], and digital apps for eHealth [42-44]. As SnehAI represents an AI-driven chatbot accessible through the Facebook Messenger mobile app, our literature review focused on discussions regarding the affordances of all relevant technologies and those pertaining to health promotion and education to set the conceptual foundation for our analysis.

As the internet proliferated, reviews about its affordances emphasized its accessibility through broadband networks. Increasingly available anywhere, anytime, and in multiple media modalities (eg, text, audio, and visual), the internet embodied the novel nonlinearity of user experience beyond time and space enabled by hypertexts and search engines [38,39]. Key affordances of mobile devices include portability of smartphones, information exchange with multiple others across multiple channels with increased frequency and directness, surveillance risks tied to locational presence, and multimedia content production and sharing [40]. The key affordances of social media platforms include the permanent nature of web-based expressions that are digitally recorded and archived; the possibility for users to re-edit and recraft their messages to manage their self-presentation; the possibility for the digital content to be searched, replicated, and disseminated to a mass audience; and the increased visibility of users' personal profiles, social connections, and web-based behaviors [41,45,46].

Although AI and chatbots have been around for decades, they have gained in currency with recent developments in cloud computing and machine learning to leverage big data [17,47]. The key affordances of AI-driven chatbots include receiving messages such as real-time status updates and aggregated information from different sources; setting options for automated reminders, nudges, and other triggers for user engagement; enabling users to query information based on personal interests; and enriching messages by having the content processed and enhanced with hyperlinks, emojis, graphics interchange formats (GIFs), and other audio-visual effects [17]. Although exploratory studies of AI-driven chatbots are scarce, their promise to address, circumvent, and overcome the taboo nature of SRH and engage young people is unmistakable [22,47].

Research Gaps and Questions

For the most part, scholars and practitioners agree on the *existence* of affordances as specific *functions* of materials or artifacts (in our case, technologies) and the necessary *recognition* of such affordances for their *realization* to take place [17,30,44]. However, there were several gaps. First, many phrases, such as *social affordances*, *cognitive affordances*, *emotional affordances*, and *therapeutic affordances*, have become popular in the literature that are in fact inferences or implications of the functional affordances [39,42,43]. Therefore, any analysis of affordances should fundamentally be anchored to technological functionalities. The categorization of social, cognitive, emotional, and therapeutic affordances should be secondary. Making such conceptual distinctions is important.

Second, most of the literature on affordances has a positive bias and is most often limited to their potential for achieving personal goals or public good without considering concerns or risks. Painting a rosy picture of a rapidly changing media landscape is dangerous. Similar to all previous information and communication technologies, smart and connected mobile devices and social media platforms work like a double-edged sword. It is our ethical obligation, in this digital and network society, to reduce the prejudice of technological determinism [31] and make deliberate efforts to address and account for the negative affordances that can cause detrimental effects [48].

Third, AI chatbot design and research are still in their infancy [14]. As technologies advance to highly interactive user interface with realistic digital representations, it is unclear how their connection to certain contextual information, such as a background story, will affect user appeal and engagement [16,49-51]. Such technological advantages tend to serve privileged populations better and can implicitly embody prejudices against those who are vulnerable and marginalized [52].

Finally, although all the technological artifacts are embedded in some kind of regulatory system, be it legal or social, the role of these regulatory systems is rarely included in the discussion of affordances. Although we focus on what algorisms would enable or prohibit digital citizens from doing (or not doing) with their smart devices, affordances of AI operate in the space of public discourses that can affect individual autonomy and privacy [53,54].

On the basis of the above literature review and identified gaps, we investigated the SnehAI chatbot as an AI-driven conversational agent and a digital tool for sex education and communication about SRH, nested in the Facebook Messenger mobile app. By applying the theory of affordances as a high-level conceptual framework, by connecting and

synthesizing all relevant affordances from the literature, and by analyzing both qualitative insights and unobtrusive quantitative behavioral data, we answered the following research questions:

1. Research question 1: What are the functional affordances of SnehAI for promoting conversations around SRH?
2. Research question 2: How are users engaging with SnehAI in their conversations?

Methods

Study Context

SnehAI was first launched in April 2019. Although its current version is a stand-alone AI chatbot, its original idea and predecessor were deeply rooted in the Population Foundation of India's innovative initiative called *Main Kuch Bhi Kar Sakti Hoon* (*I, A Woman, Can Achieve Anything*). This initiative used a 360° approach, especially the *transmedia edutainment* social and behavior change communication strategy [55,56], coordinated across multiple media platforms to challenge deeply entrenched regressive gender norms and to advocate for women's empowerment. Powerful stories were told through the dramatic journey and positive role modeling of protagonist Dr Sneha, who leaves behind a lucrative medical practice in Mumbai and returns to her home village Pratappur after her sister's death from a forced abortion. She then stays on to tackle multiple social ills—child marriage, sex selection in favor of male offspring, violence against women and girls, and many other manifestations of gender inequality [57].

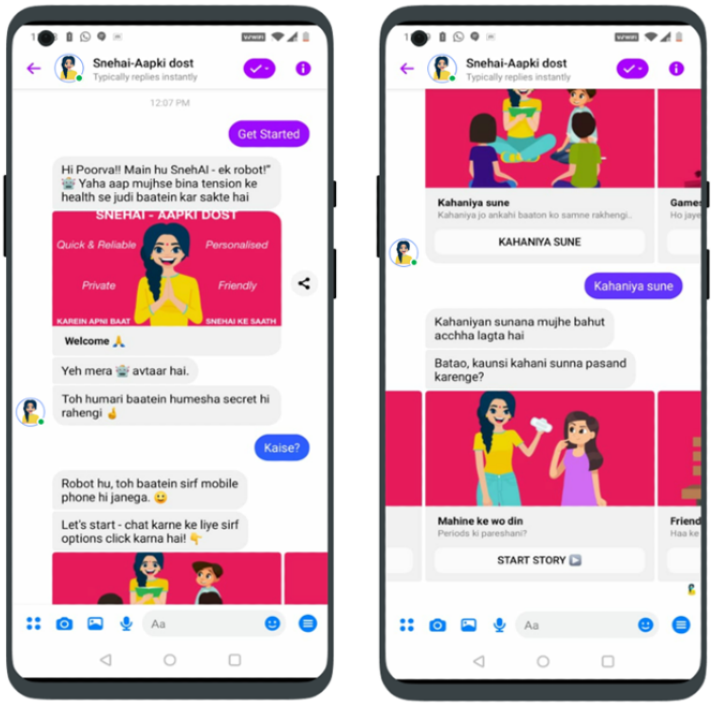
From March 2014 to September 2019, 3 seasons of *Main Kuch Bhi Kar Sakti Hoon* were broadcast on the Indian national television network Doordarshan, hundreds of All India Radio stations, and the Mobile Vaani community [58,59]. *Main Kuch Bhi Kar Sakti Hoon* was the first to use an interactive voice response system for large-scale real-time audience engagement, creating a *voicebook* as a transmedia extension that allowed millions to interact with curated content, answer questions, and share personal opinions and actions inspired by the characters in the program [58]. Several miniseries were created out of the television content, extending the storyline on *Main Kuch Bhi Kar Sakti Hoon*'s Facebook page and YouTube channel [57,60]. In particular, its Facebook page was popular among young people and male users according to Insights Analytics in October 2019 (Multimedia Appendix 1).

In April 2019, during season 3, *Main Kuch Bhi Kar Sakti Hoon* launched a new transmedia extension through Facebook Messenger: a chatbot named after Dr Sneha, *SnehAI* (for Sneha AI). SnehAI was created by the Population Foundation of India in close partnership with the UK-based innovative technology company AI for Good. This AI chatbot was purposefully designed to extend Dr Sneha's media personality as *a trusted friend* and open a safe, nonjudgmental, and private channel to engage the audience (especially young people) in India to talk and learn about SRH. SnehAI version 1.0 used a built-in decision tree with predetermined rules and quick replies to help users navigate through the visual menus [25,26]. It was cocreated with invaluable inputs from 84 adolescents and 19 adults to ensure a friendly tone of voice with familiar colloquial expressions that, in fact, SnehAI became the world's first ever Hinglish AI chatbot developed for social and behavioral change [25,26]. It was promoted through *Main Kuch Bhi Kar Sakti Hoon*'s website, interactive voice response system, Facebook page, and other channels. Building on previous efforts, this paper focuses on the current version of the AI chatbot, SnehAI version 2.0.

SnehAI Chatbot

In April 2020, SnehAI (version 2.0) was launched with a natural language processing (NLP) platform and better content flow to make SnehAI more intelligent in her conversations with users [26]. It is required that an individual be a user of Facebook Messenger to chat with SnehAI (Figure 1).

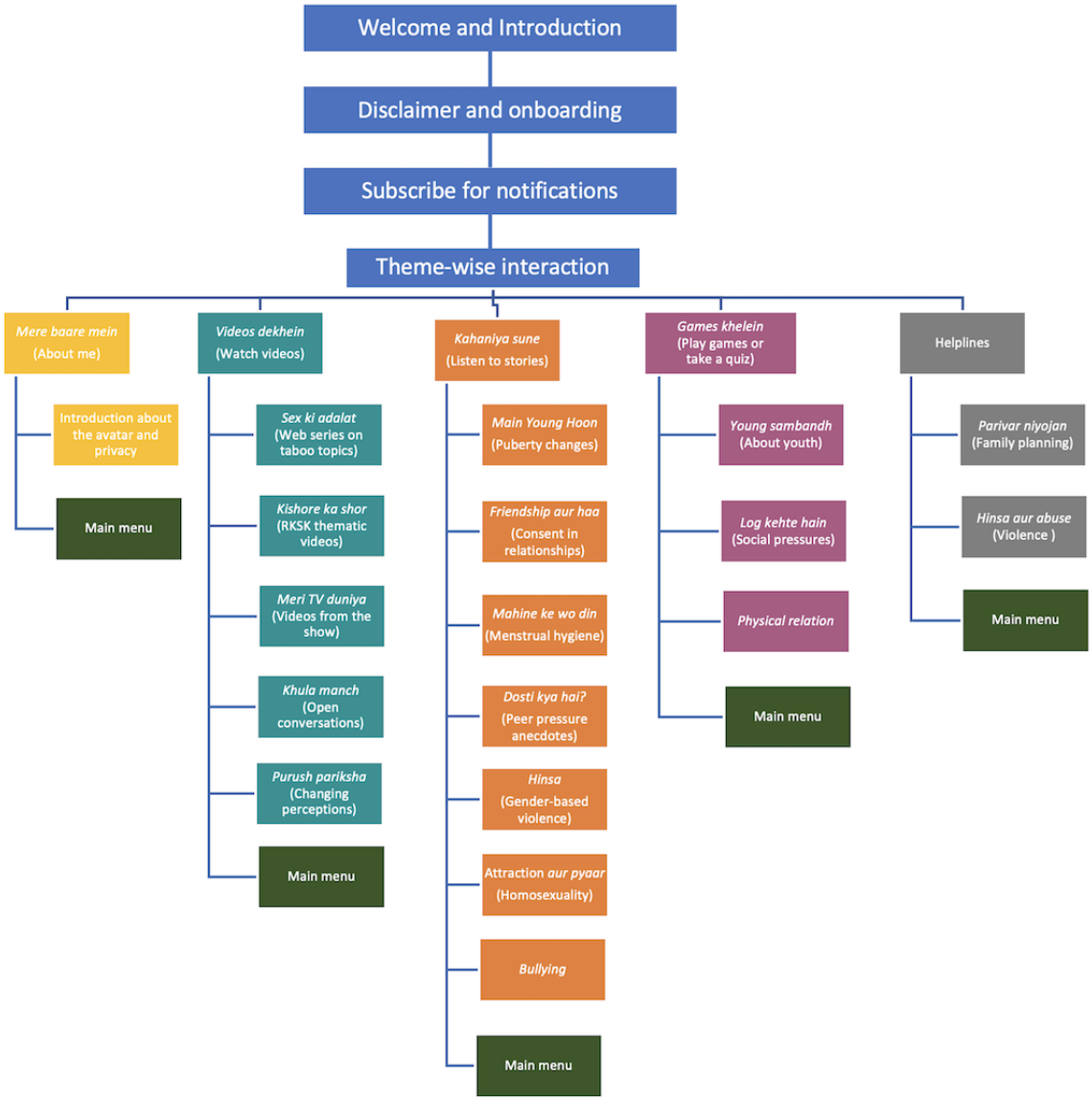
Figure 1. SnehAI chatbot user interface on Facebook Messenger.



The conversation between a user and SnehAI starts with a warm welcoming message. The chatbot introduces itself as Dr Sneha. She mentions that she works in a hospital and is involved in other development projects in the village. After introduction, she asks the user to choose one of the available options to ensure that the user understands that he or she is interacting with a chatbot rather than a human being. After that, the users are invited to subscribe to notifications from SnehAI. Then they can select and interact with a range of content presented on the main menu. Depending on their choices, the information may be displayed as hyperlinked texts, images, or short videos. Furthermore, these theme-wise interactions are organized into five content categories (Figure 2):

1. *Mere baare mein* (About me): this is an introduction to the app. SnehAI is introduced as an avatar, and users are also given the option to review the privacy policy. In the chatbot, this option appears at the end, as the goal is to promote health-related communication.
2. *Videos dekhein* (Watch videos): this option allows the user to watch short videos from 3 interrelated programs. *Sex ki adalat* includes 4 fictional court cases debating about virginity, masturbation, menstruation, and pornography; *Kishor ka shor* offers conversations based on the government-endorsed national adolescent health program RKSK themes featured in *Main Kuch Bhi Kar Sakti Hoon* season 2 regarding obesity, child marriage, and cigarette smoking; *Meri TV duniya* includes trailers and exclusives from *Main Kuch Bhi Kar Sakti Hoon* season 1, 2, and 3. *Khula manch* offers conversations on mental and physical health; and *Purush pariksha* challenges beliefs and pushes viewers to question orthodox ideas.
3. *Kahaniyan sune* (Listen to stories): this option allows the user to engage with specific storylines featuring 7 young female and male characters in *Main Kuch Bhi Kar Sakti Hoon*. These stories include Naveen's contemplation of adolescent attraction and curiosity and confusion about puberty (*Main young hoon*), Pinky's story about menstrual hygiene (*Mahine ke vo din*), Shama and Aman's relationship and their beliefs about consent (*Friendship aur haa*), Sanjay's issues with peer pressure (*Dosti kya hai?*), Preeta's survival of an acid attack for being a competent female football player (*Hinsa*), Bunty's homosexuality (*Attraction aur pyaar*), and Aman's concerns about bullying (*bullying*).
4. *Games khelein* (Play games or take a quiz): this option allows the user to play quiz games on three themes: *Young sambandh* has questions about adolescent development, *Physical relation* has questions about family planning, and *Log kehte hain* has questions about myths and misconceptions about SRH.
5. Helplines: this option directs users to 2 helplines. One helpline is on SRH through the national toll-free number of the Jansankhya Sthirata Kosh endorsed by the Ministry of Health and Family Welfare. It provides callers with information and counseling on family planning and reproductive health issues, including access and use of contraceptives. The other helpline option links users to a national helpline on gender-based violence for the counseling and reporting of incidents.

Figure 2. SnehAI chatbot content flowchart. RKSK: Rashtriya Kishor Swasthya Karyakram.



In addition to the improved content structure and flow, a critical element of SnehAI (version 2.0) is the use of an NLP system to process and respond to user-entered free-text queries. This NLP system first matches the free texts with prefilled regular expressions (RegEx in short as commonly used by programmers), and if a match is found, corresponding replies are executed. If no matches are found, the NLP system activates a Microsoft Language Understanding (LUIS) conversational AI service and uses it to apply custom machine learning intelligence in client apps (such as AI chatbots, such as SnehAI) to process user input and provide appropriate response [61]. The Azure LUIS app, commonly used in AI chatbots, receives user input in the form of conversational and natural language texts, treats them as *user utterances*, processes the information by extracting *keywords* and predicting *user intentions*, and then uses the JSON response to make decisions about how to fulfill the user's requests [61].

Instrumental Case Study

The case study approach is particularly useful for exploring a phenomenon in its natural and real-life context and allows the development of a more in-depth, multifaceted, and holistic understanding [62]. Although this method is well-recognized in disciplines such as business, law, and policy, it needs to be embraced more in global and public health research [63]. We adopted an instrumental case study approach. An *instrumental* (vs *intrinsic* or *collective*) case study focuses on 1 example to gain a broader appreciation of a phenomenon [63,64]. The case in point here is SnehAI as an exemplar of AI chatbots for promoting conversations about SRH. This research method allowed us to explore SnehAI from diverse perspectives, such as technology design, program implementation, and user engagement. We also used a mix of qualitative insights and quantitative data to triangulate our findings.

Qualitative insights were obtained from key stakeholders, with institutional representatives being the coauthors of this paper

and others listed in the acknowledgments. A total of 2 in-person group meetings, 4 virtual conferences, and numerous follow-up email discussions took place from September 2018 to December 2020. Half a dozen interim and final reports (including 2 cited in this paper [25,26]), as well as detailed meeting notes, served as foundational documents for our analysis. Quantitative data were obtained through unobtrusive chatbot user behavior tracking with user permission. The chatbot user interactions were mainly captured in 2 ways. First, for any predetermined options indicated in the SnehAI flowchart, the choices were recorded anonymously in a third-party digital analytic tool designed for chatbots on Facebook Messenger, Dashbot [65]. Conversation-specific analytics were aggregated to monitor user engagement and help improve the dynamics of user–bot interactions. Second, a user could also type questions or comments in the dialogue box at any time. These free texts were used as user input in the NLP system to detect user intention and determine the optimal response. For this project, we used the aggregated analytics data from the Population Foundation of India and its partner AI for Good, United Kingdom, gathered from the beginning of May 2020 to the end of September 2020. The data science team at AI for Good, United Kingdom, also randomly selected 15,000 free-text messages with over 20 characters and used machine learning techniques to detect user behavior patterns.

To proactively address potential pitfalls of the case study methodology [63], we adopted a widely accepted and multidisciplinary conceptual framework (ie, the theory of affordances) to guide our inquiries and investigations. We also drew clear boundaries around the specific case (ie, SnehAI 2.0) and the timeframe of user engagement data for our analysis (ie, a 5-month period shortly after the launch). Furthermore, we went through several iterations of data validation, reported discrepancies caused by technical errors, and shared detailed notes with all team members to ensure transparency in the research process.

Results

Functional Affordances of SnehAI

To answer research question 1, we compiled an extensive list of relevant affordances reviewed in the literature, evaluated them against the SnehAI app based on the qualitative insights we obtained, and distilled 15 functional affordances (see a summary in Multimedia Appendix 2 [17,33,34,38-41,45,53]). The results presented here specifically address the research gaps we identified. The affordances of a new technology should be primarily about its functionalities. Therefore, these 15 affordances of SnehAI are all based on the functionalities of the chatbot, its AI capacity through the NLP model, the Facebook Messenger mobile app it is embedded in, and the internet and wireless networks that enable its existence. We also included the role of a media personality and relevant legal and social regulatory systems in our assessment. Moreover, we considered both positive and negative sides of each affordance, with the intention of appreciating its actionable possibilities while mitigating potential concerns. Certain labels of these affordances (ie, *accessibility*, *multimodality*, *nonlinearity*, *editability*, and *scalability*) had been previously used in the literature, and some others were differently labeled although they expressed similar ideas (ie, *customizability*, *interactivity*, *visibility*, *glocalizability*, and *connectivity*). Notably, we contributed several new labels (ie, *compellability*, *queriosity*, *trackability*, *inclusivity*, and *actionability*) to represent critical affordances of SnehAI that could apply to similar AI and chatbot apps:

1. *Accessibility* allows SnehAI users to access SRH information that is accurate (with content experts' approval), trustworthy (as shared by the avatar of a trusted friend), and relatable (with messages expressed in colloquial Hinglish). The subscription option for notifications affords additional access to new information as it becomes available. However, digital inequalities may deprive underserved population groups such as women, youths, and rural residents of smartphone ownership or limit their personal use. Furthermore, users who obtain access to the chatbot can potentially be overwhelmed by new information about taboo topics. It helps that most information on SnehAI is presented as edutainment, both entertaining and educational.
2. *Multimodality* affords SnehAI to be presented through more than one sensory mode to help enhance the user experience. Although common practice in the industry is to simply use plain text in a chat box, SnehAI uses rich media that includes text, audio, and visual modalities, and popular presentation forms, such as GIFs, emojis, and short videos. What SnehAI users see and hear about gender norms may be dramatically different from what is covered in the mainstream media, which could potentially be a source of cognitive dissonance, yet necessary for raising social consciousness about gender inequality [57-59].
3. *Nonlinearity* allows SnehAI to provide curated content through clickable visual menu options instead of a prescribed sequence. In addition, users can enter their queries through free-text messages at any time apart from when they access branched content categories. Thus, SnehAI allows users to take control of their experience as more of a choose-your-own-adventure type of journey. On the flip side, such an adventurous and nonlinear trajectory can also be distracting or confusing for some users who may lose track of a thought as they shift from one content category to another. Currently, there is no separate search bar available in the user interface that allows users to conduct a keyword search from anywhere.
4. *Compellability* allows SnehAI to engage users in compelling ways. The casual, friendly, yet culturally appropriate–looking avatar of SnehAI is rich in visual appeal for Indian adolescents and young adults. Paid Facebook promotions can help attract new users and boost user retention. Furthermore, the triggers and prompts on Facebook Messenger can also attract user attention and provide cues to action. Although no nudges exist presently when a user is inactive for a prolonged time, nudges can be annoying if their intentions are unclear to users [17].
5. *Queriosity* allows SnehAI users to seek answers to their queries based on personal curiosities. For users to be able to enter free-text messages is in stark contrast with the

conventional method of preaching to the choir of users. In particular, young people learn better when their genuine curiosity is encouraged without judgment [55,56]. The analysis of the 15,000 sampled queries by our data science team indicated that "SnehAI opened up a safe space for users to ask 'stupid' or 'embarrassing' questions about SRH without worrying about what others think." The current version only allows users to query through manually typed text messages. In the future, other query methods may be included. For instance, a search engine may be helpful when the curated content database expands so that users can find and explore the most relevant branched content category. Providing voice input and output capability (as in Siri and Alexa) can be especially useful for users with low levels of literacy.

6. *Editability* allows SnehAI users the ability to craft a message to their satisfaction before sending it out to the chatbot. Thus, users have full control over how they present themselves in this private space. Again, the analysis of 15,000 sampled queries by the data science team showed that most user behaviors were respectful. "For instance, many treated the chatbot as if it was Dr. Sneha, addressing her as Dr, Ma'am ji [respected Ma'am], and didi [older sister]." On the other hand, once a free-text message is sent out, there is no way to revise or recall it, even to correct unintentional mistakes.
7. *Visibility* allows SnehAI to automatically save all messages exchanged between the user and the chatbot and keep them permanently visible in the chat history. Such permanence in visibility can be a convenient information repository over time but also a potential vulnerability, as anyone who has access to the user account can retrieve and review private messages. SnehAI is proud of its strict adherence to the international General Data Protection Regulation on data privacy with a built-in user option to review its detailed privacy policy, including user permission for recording or deleting data.
8. *Interactivity* allows SnehAI to provide immediate feedback on user requests. This allows the chatbot user to simulate real-time interpersonal interaction through turn-taking with verbal expressions and nonverbal cues—through emojis. Therefore, each conversational session can flow like text messaging with a real person. SnehAI is prompt and reliable in responding. However, the free-text message queries by a user are entirely dependent on the quality of the response through the NLP model. When the user intention is incorrectly decoded, the chatbot response is correspondingly off-track. Thus, the quality of conversations with the chatbot is determined by the intelligence of the NLP model in detecting communicative nuances and minimizing bias.
9. *Customizability* allows SnehAI users to interact with curated content and customize their queries based on personal interests, even though there is no option for users to customize their chatbot menus or filter content. Nevertheless, SnehAI uses a hybrid of prefilled RegEx and the LUIS conversational AI client app in its NLP system to customize the chatbot's response to user queries. In particular, the NLP uses *machine learning features*, such as a phrase list, to detect *user intentions* at various junctions of the user journey [66]. For example, *small talk intents* cover the trivial chit-chat with phrases such as *how are you* and *thank you*. However, a unique aspect of SnehAI is that the NLP model is designed to effectively detect *core intentions* related to topics such as safe sex, reproductive health, and family planning choices. These *core intentions* are identified through keywords that serve as substitutes for *sex*, accounting for the corresponding colloquial expressions by young people and their spellings and misspellings in Hinglish. The NLP system used in SnehAI version 2.0 was initially trained based on insights from the previous version and is being retrained regularly. A part of the LUIS global network, SnehAI contributes Hinglish conversational messages to the larger pool to improve machine learning intelligence over time. It also leverages its growing database for SnehAI NLP training. A more detailed illustration of the LUIS NLP model in SnehAI is provided in Multimedia Appendix 3 within a linked video in Multimedia Appendix 4.
10. *Trackability* allows SnehAI the capability to unobtrusively track certain users' personal information and interactive behaviors with the chatbot. Because of privacy concerns, Facebook is strict about the kind of personal information recorded in third-party apps. For SnehAI, version 2.0, a formal application was submitted, reviewed, and approved by Facebook to record user's gender in the Dashbot analytics tracking. Other key performance indicators such as important clickable reactions were also recorded and reported in the *User Engagement With SnehAI* section in *Results*. In addition, inappropriate and abusive user behaviors were also monitored, flagged, and handled to keep SnehAI a safe and civil space. No matter how elaborate the efforts to protect user privacy, some users (particularly those with low digital media literacy) may not fully understand what kind of personal information is being tracked at the back end.
11. *Scalability* allows SnehAI to achieve large-scale user reach by leveraging the popularity and massive user base of Facebook in India, especially among adolescents and young adults. All the information on SnehAI is digitally produced and can be easily replicated for dissemination. As a stand-alone mobile app, SnehAI can be an effective digital tool for frontline health workers to share with young men and women. Harnessing the existing health care and field-based health education networks can easily spread SRH resources on SnehAI across tens of millions of users, with the probability of some creative content going viral. Currently there is no option in SnehAI for users to recommend the chatbot to other Facebook friends. In addition, no aggregate user engagement information is shared in the chat function to further grow the SnehAI user base.
12. *Glocalizability* allows SnehAI to reach the youth living in India and around the globe. The intentional use of colloquial Hinglish makes the AI chatbot particularly approachable and friendly for young people. Regardless of where a user lives, there is a certain level of universality with respect to curiosity, challenges, and experiences related to SRH issues. It is commendable that SnehAI, version 2.0, offers

information on 2 nationwide helplines. Unfortunately, no location-based content tailoring is available at this time to contextualize the SRH information for individual users, nor is there a capability to facilitate direct connections with existing local health and social services.

13. *Inclusivity* affords SnehAI to be welcoming of users with diverse backgrounds, especially underserved population groups such as women, youths, and rural residents. Once SnehAI can be accessed, the service is free. In addition, AI chatbots are becoming increasingly capable of processing human voice as user input, such as Siri and Alexa. Rather than depending on the users' reading and writing skills, a voice-based option can benefit people with low literacy and make SnehAI even more inclusive. Meanwhile, the chatbot presents some privacy risks for users with low digital literacy. Imagine a worrisome parent finding a teenager asking about sexually transmitted infections or a controlling husband finding his wife learning about injectable contraception. With time, other functions may be incorporated to provide additional education and protection to the most vulnerable users.
14. *Connectivity* allows SnehAI meaningful opportunities for parasocial interactions to occur between a user and the chatbot via Dr Sneha's avatar. Parasocial interactions typically occur in pseudorelationships between a media user and a media personality [67]. For SnehAI users, having someone they can trust to talk about SRH is much better than having no one. According to our data science team, "Many opened up and shared deeply personal issues. Some were keen in knowing more about Dr. Sneha's whereabouts, wanting to talk in person, and even asked for her WhatsApp number." In the current version, options are not yet available to safely connect a user with other trusted real-life friends through Facebook Messenger. There is also no direct connection to link one user's queries with others' queries to securely facilitate peer-to-peer or social learning.
15. *Actionability* affords SnehAI users to take action as the direct result of their conversations with the chatbot. For example, they can share in-person SRH information that they have learned with their trusted peers. Such private and casual information sharing is, in fact, how many people learn about SRH in the first place. They can also call national helplines to seek professional consultation for themselves or others. Currently, there is no possibility of directly calling through the AI chatbot app or appointment scheduling capabilities with local health or social services. However, with the rapid development of AI technologies such as Google Assistant, more affordances can and will become available for users in due time.

User Engagement With SnehAI

To answer research question 2, we examined the aggregate Dashbot analytics data on SnehAI user behaviors. Over a 5-month period, 8,170,879 messages were exchanged between SnehAI and 135,263 unique chatbot users, including 5,100,449 (62.42%) outgoing messages from the chatbot to the users and 3,070,430 (37.58%) incoming messages from the users to the chatbot. The ratios between outgoing and incoming messages were consistently around 60:40 over time, with the exception of repeated outgoing messages in July (such as technical errors, which are marked with asterisks in Figure 3). Specifically, incoming messages from the users to the chatbot were sent through free texts (1,599,339/2,878,908, 52.09%); clickable reactions (1,279,569/2,878,908, 41.67%); or GIFs, audios, images, and videos (191,522/2,878,908, 6.24%).

The average user engagement with SnehAI was 1.9 sessions, 7.6 minutes, and 56.2 messages exchanged (Figure 4). The time spent by top users increased from 2 to 3 hours in the first 3 months to 7 hours in 14 sessions in August and 14 hours in 47 sessions in September.

Figure 3. SnehAI chatbot outgoing versus incoming message ratios over time.

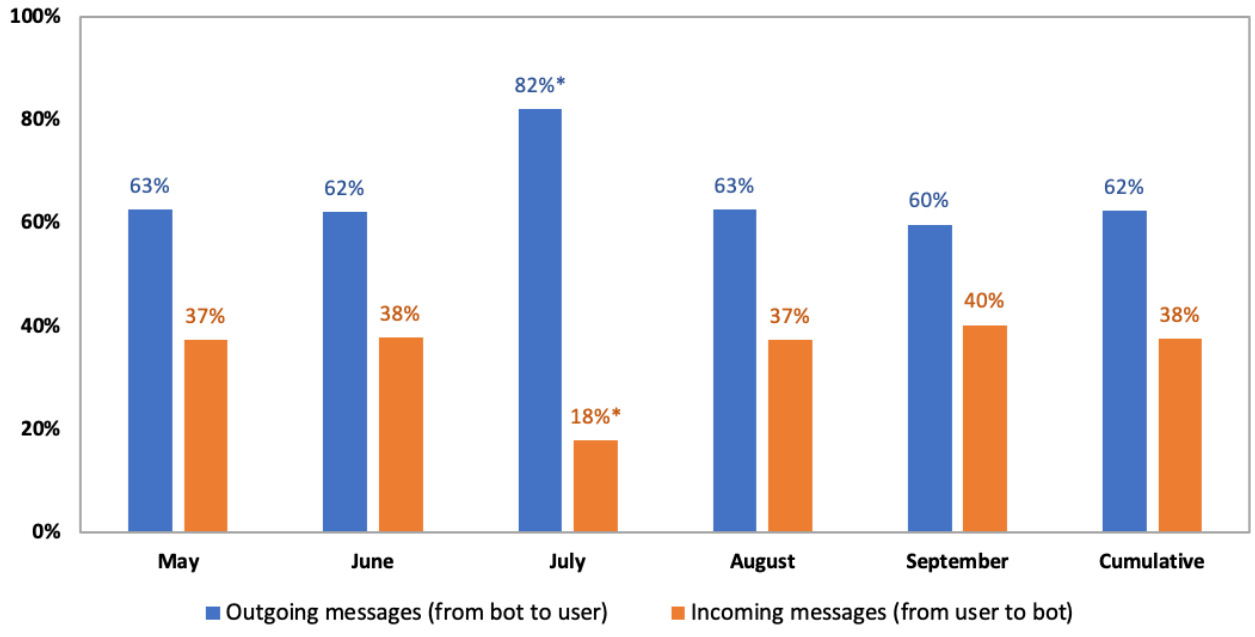
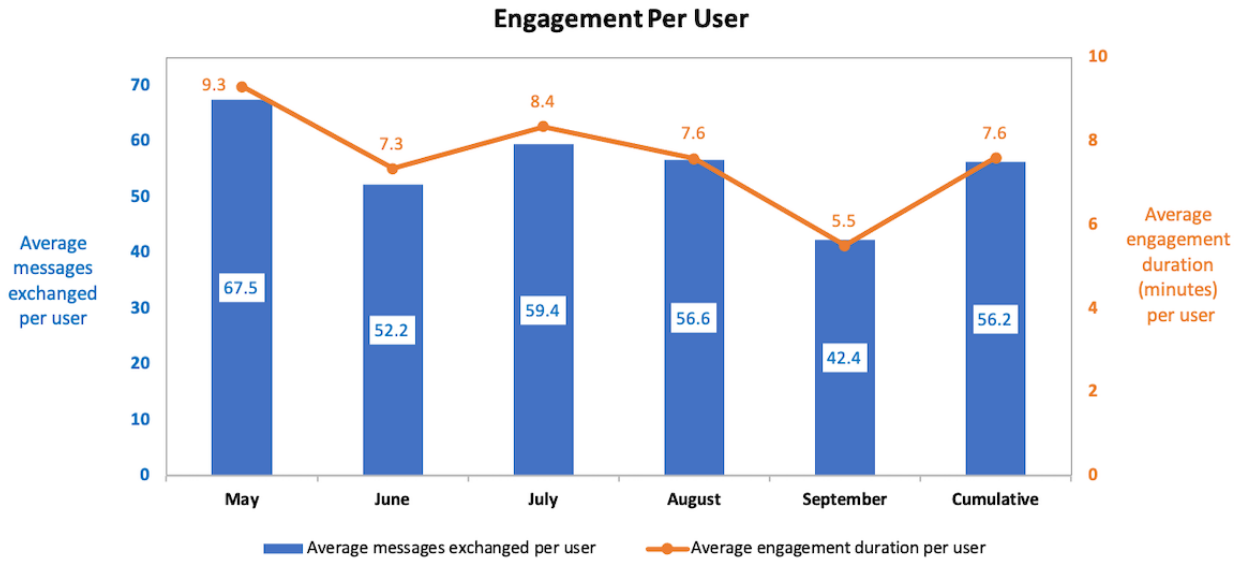


Figure 4. SnehAI chatbot average user engagement over time.



User engagement was also tracked through the content categories, including both guided flows using clickable reactions and the handling of free-text messages using NLP. Although multiple messages could have been exchanged with SnehAI each time a user engaged with a particular content category, the overtime trends of engagement frequencies across different content categories provided empirical evidence of user response and preference. From onboarding, to learning about Sneha, to its privacy policy, and its main menu to the videos, stories, games, helplines, and query responses through NLP, the chatbot users traversed across these content areas for 1,430,416 times over the course of 5 months. Approximately half (705,305/1,430,416, 49.31%) of these interactions were about the chatbot responding to the user queries, including small talks and any questions related to themes of health communication (Figure 5). The next highest frequency content category was onboarding (257,042/1,430,416, 17.97%), about SnehAI (115,131/1,430,416, 8.05%), privacy policy (96,305/1,430,416, 6.73%), main menu (83,692/1,430,416, 5.85%), helplines (71,211/1,430,416, 4.98%), stories (61,582/1,430,416, 4.31%), games (20,897/1,430,416, 1.46%), and videos (19,251/1,430,416, 1.35%). These trends in content engagement distribution were relatively consistent (with occasional fluctuations).

Our analytics tracking data showed a count of 99,936 typed text messages from the users that SnehAI handled and responded through the NLP system with queries related to the six topical themes (Figure 6): safe sex practices, such as consent, frequency of sexual intercourse, oral and anal sex, impact of other health ailments on sex life, and unplanned pregnancy (57,158/99,936, 57.19%); choice of family planning methods, such as male and female condoms, oral contraceptive pills, intrauterine devices, injectables, and SRH-related themes, such as abortion, sexual intercourse during and after pregnancy, polycystic ovarian disease, and infertility issues (6287/99,936, 6.29%); female reproductive health concerning menstruation (eg, regularity, pain, discharge, and spotting), virginity, and premarital sex (13,965/99,936, 13.97%); adolescent sexual health issues, such as nightfall, masturbation, pornography, sexual stamina, erectile dysfunction, and STIs (15,160/99,936, 15.17%); adolescent mental health issues regarding peer pressure and bullying (2343/99,936, 2.34%); and nutrition and social determinants of health, such as child marriage and gender equality (4623/99,936, 4.63%).

With Facebook's approval, we accessed the SnehAI chatbot user's gender-disaggregated data. What we discovered was an extreme gender gap (Figure 7): among the unique chatbot users over this 5-month period, 93% (125,795/135,263) were male, 6.8% (9198/135,263) were female, and 0.2% (270/135,263) unknown. This gender ratio was disproportionally skewed toward male users compared with that in the United Nations Development Programme report on gender distribution of internet users (29% female) and Facebook users (22% female) in India [12], as well as with that of the general population (48% female) according to census data [2].

Figure 5. SnehAI chatbot content engagement distribution over time.

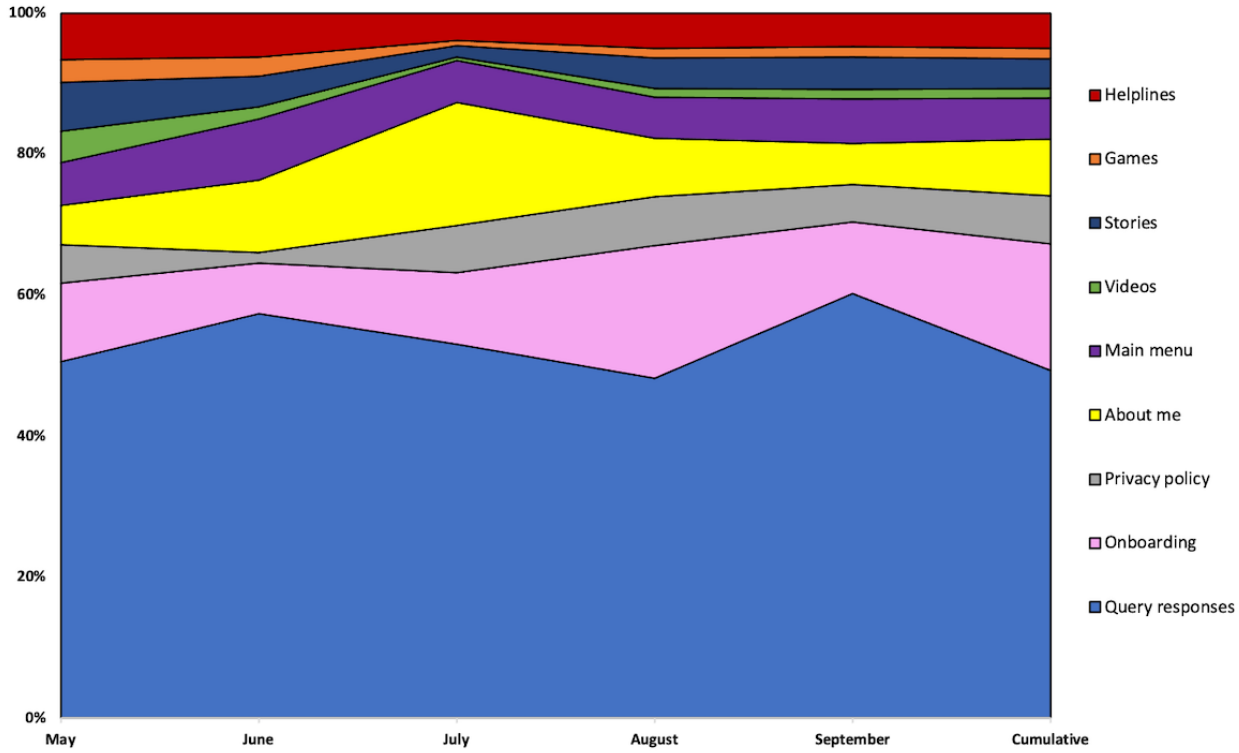


Figure 6. SnehAI chatbot free-text queries content distribution over time.

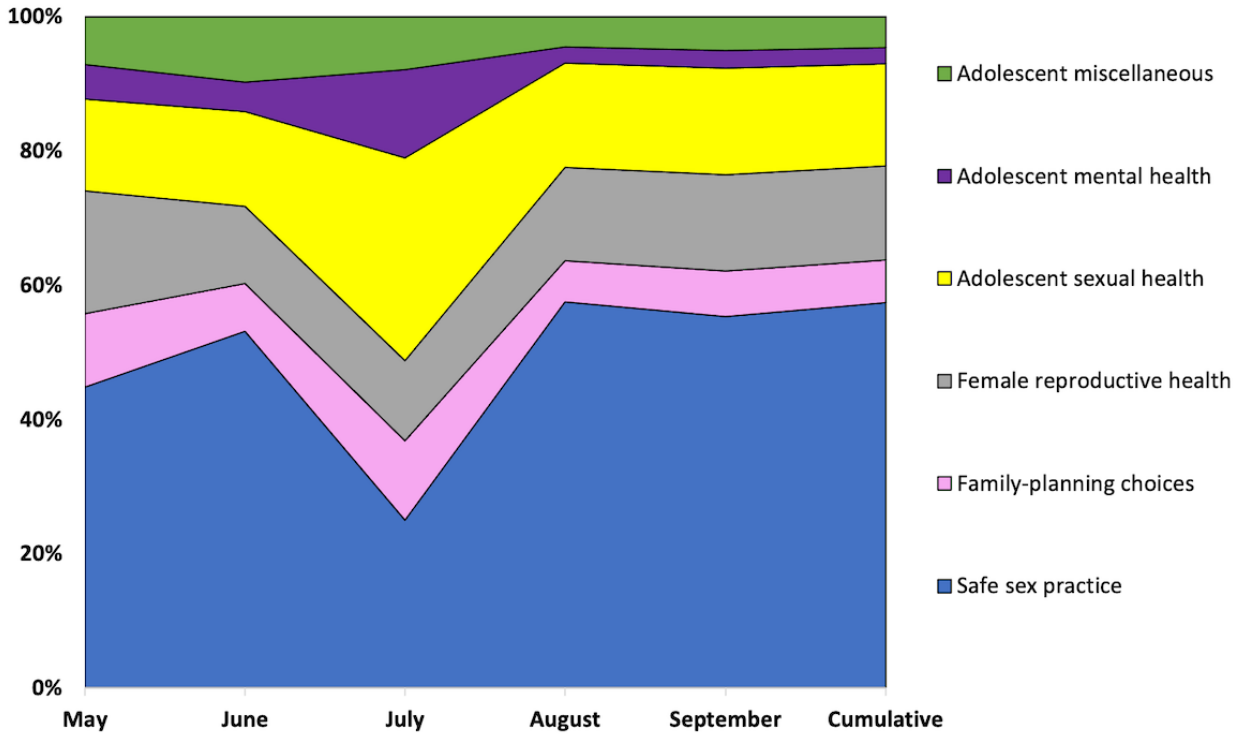
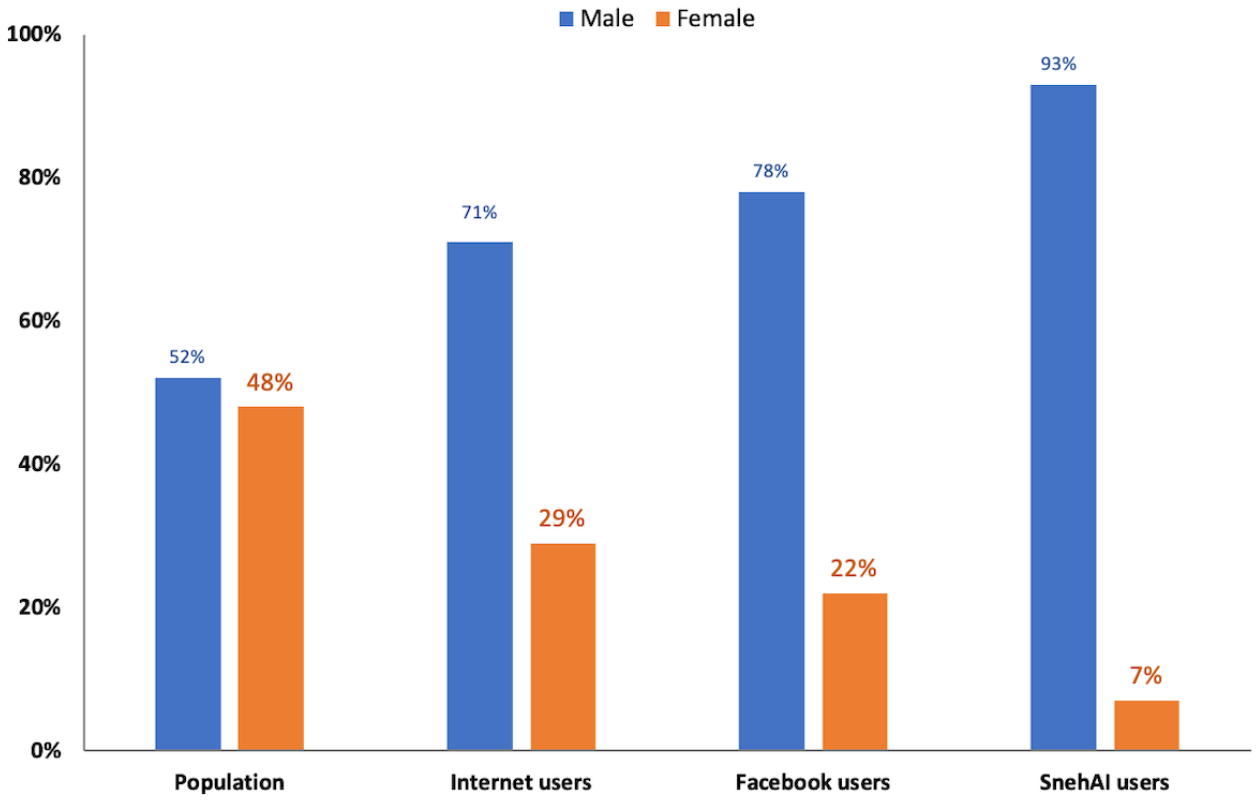


Figure 7. Gender gap in SnehAI users compared with internet and Facebook users in India.



Even worse, based on the 15,000 sampled queries, our data science team found "a behavioral pattern with female users showing significantly lower self-confidence in their conversations with SnehAI." This might be a function of female users facing gender disparities in mobile device ownership and low digital media literacy in India.

Furthermore, they are likely less comfortable about openly discussing SRH and may hold higher levels of privacy concerns for digital abuse and sexual exploitation. In contrast, there were incidents where male users requested SnehAI to set up a girl for them, and some even engaged in obscene chats with her expressing intimacy through sexting, asking for porn videos, and using abusive language. A few questions also showed the darkest side of sexual abuse, with male users inquiring which tablets they could give girls if they were unwilling to have sex with them, with the intention of overriding consensual sex.

Discussion

Principal Findings

SnehAI pioneered the use of an AI chatbot to engage users, especially young men, in conversations about SRH. A total of 8.2 million messages were exchanged in a 5-month period, with almost half of the incoming user messages being free-text queries comprising personal questions and concerns about SRH. Overall, the SnehAI avatar based on Dr Sneha—the protagonist in *Main Kuch Bhi Kar Sakti Hoon* television show—presented itself as a trustworthy friend and mentor. Furthermore, the curated content from this gender equality and family planning initiative was found to be both entertaining and educational. Moreover, the NLP system worked smoothly and effectively to personalize the chatbot response and optimize user experience.

Using the theory by Gibson to guide our research inquiries and building on the existing literature, we distilled 15 functional affordances of SnehAI. The results of our instrumental case study provide evidence that SnehAI offers a safe space for users to talk about sensitive SRH topics, seek and obtain accurate information, access services locally through helplines, and seek personal counseling:

1. *Accessibility*: SnehAI provided a digital platform for Indian youth to access highly accurate and credible information about SRH in a safe, nonjudgmental space.
2. *Multimodality*: SnehAI used popular media modalities such as videos, GIFs, and emojis to enhance engagement.
3. *Nonlinearity*: SnehAI offered curated content in branched categories that invited nonlinear navigation.
4. *Compellability*: SnehAI compelled the users to open up through her relatable and approachable avatar image and tone.
5. *Queriosity*: SnehAI welcomed queries that encouraged user curiosity about SRH topics.
6. *Editability*: SnehAI enabled users to reword, redraft, and rescript how they showed up in the digital space.
7. *Visibility*: SnehAI enabled users to privately and securely save chat history, making it visible at any point.
8. *Interactivity*: SnehAI simulated conversational turn-taking and immediate verbal and nonverbal feedback.

9. *Customizability*: SnehAI enabled users to navigate the curated content according to their interests and used its machine learning capability to customize responses.
10. *Trackability*: SnehAI tracked users' web-based behaviors through their clickable actions and self-generated content.
11. *Scalability*: SnehAI holds tremendous potential for scaling up through various adolescent health education programs, networks of frontline health workers, and diverse service providers.
12. *Glocalizability*: SnehAI enabled linkages between young people and service providers in their local areas.
13. *Inclusivity*: SnehAI served as a free digital consultation repository for all users.
14. *Connectivity*: SnehAI created the conditions for users to connect parasocially with a respected and trustworthy avatar at a deeply personal level.
15. *Actionability*: SnehAI used the power of transmedia storytelling to inspire users to take concrete actions to seek help for themselves and their peers.

Meanwhile, we also discovered an extreme gender gap among the many potential disparities. The patriarchal restrictions on women's autonomy, mobility, and self-expression in India are vividly reflected in SnehAI user analytics data. This disparity in ratio is a risk and can introduce gender bias in the AI chatbot's NLP model. Our team is currently adapting field-based promotional strategies to include more female users. Other specific opportunities for improvement based on our results include the following:

1. A more intelligent NLP model coupled with targeted behavior change strategies focusing on awareness building and empowerment of women and girls to reduce gender bias.
2. A system setting of reminders and nudges to re-engage idling and inactive users.
3. An option to edit or recall a text message after sending it out to the chatbot or to use voice-based input and output to include users with low literacy.
4. A possibility to customize the menus, filter contents, and search information in the chatbot when the digital artifacts propagate to a larger scale.
5. A feature that allows the user to invite other trusted friends to use the chatbot and leverage anonymous and aggregated user data to boost user participation and peer-to-peer learning.
6. A mechanism to educate the most vulnerable users on how to protect their privacy, safely participate in web-based activities, and advocate for their basic human rights.
7. A network that can directly connect the chatbot users to their local health and social services, for instance, enabling helpline dialing and scheduling appointments through the chatbot.

Practical Implications

Our findings hold important policy and programmatic implications for health informatics, especially for the design of user-centered and AI-driven interventions. Understanding both positive and negative aspects of the affordances of AI chatbots such as SnehAI can help inform future endeavors to reinforce and amplify the positives and minimize the negatives when designing, implementing, and studying these technologies. The 15 affordances of SnehAI can serve as a comprehensive checklist for similar apps. It helped our research team to pinpoint the unique attributes and strengths of SnehAI and clarify the next steps for future improvements. They can also help chatbot design and analytics teams to select key performance indicators to track over time and monitor user behavioral patterns to ensure the safety of the environment and improve user engagement. This case study also described the NLP model used in SnehAI, version 2.0, to customize the interactions between an AI chatbot and a user. NLP helps optimize the conversation dynamics and personalize the user's journey, thus getting us closer to accomplishing the United Nations sustainable development goals, focusing on good health and well-being, gender equality, and quality education.

In addition, by covering all relevant aspects of the SnehAI chatbot, we learned that when treating technologies as artifacts, their affordances should include not only the specific app itself. What is equally, if not more, important is to include the information systems and communication networks an app is nested in, as all of them are interconnected and interdependent. SnehAI could not be functional without Facebook Messenger, access to mobile devices, or mobile connections to the internet.

Furthermore, studies on technological affordances should routinely include attention to legal and social regulatory systems that are usually invisible to the public. This is particularly pertinent when a citizenry is divided along the lines of prejudice and hierarchy, such as class, gender, and race. The rapid development of AI technologies will continue to challenge innovators, marketers, and users with moral dilemmas. Our team's experience working directly with the intended users of SnehAI to respect the circumstances of their living conditions while facilitating meaningful social and behavioral change has demonstrated great value in cocreation and collaboration.

Theoretical Implications

Our case study on SnehAI demonstrated both the robustness and adaptability of Gibson's theory of affordances, some 40 years after it was proposed. At the same time, our investigation of SnehAI also sheds light on research gaps in the current literature on technology affordances and offers new insights to expand this theoretical framework. Only when there is a deeper understanding of the design of a user interface and its operating system behind the screen (similar to what our research team did) that it becomes possible to evaluate new technology's affordances—through their existence, recognition, and realization. Researchers should return to the basic functionalities before jumping to conclusions about other types of social, cognitive, emotional, and therapeutic inferences.

Second, there is a utopian bias in affordances in new media studies. It is important to balance technological determinism and social constructivism when examining the social shaping and consequences of emerging technologies. The identification of affordances by nature is a function of one's individual capability. An affordance is permanent, but its perception is dependent on the individual's culture, history, effort, and necessity. Consequently, it is critical for users to be at the front

and center of the design. SnehAI was developed keeping the user in mind, as the path of affordances perceived by users in India would be unique to them. Affordances incorporate the microbehaviors exhibited by users, making it imperative that there is a balance between actionable possibilities that are both positive and negative and to proactively address potential risks and concerns, especially when the intention is to better serve the underprivileged populations.

Finally, the human–AI interactions that we observed through SnehAI were intriguing. This can be viewed as a contemporary version of what Horton and Wohl [67] called *parasocial interaction* between mass media users and media personalities on radio or television. It can also be viewed as a variation of *player–avatar interaction* in game research [68]. The narrative connection to SnehAI and its friendly avatar in a private conversational setting offers several opportunities for theoretical development in human–computer interaction and computer-mediated communication.

Limitations and Future Research

Our instrumental case study did not include any primary data collection from SnehAI users. This is a necessary next step. Although *Main Kuch Bhi Kar Sakti Hoon* Facebook page analytics provided useful proxies for Facebook users, direct contact with the AI chatbot users themselves using surveys would be immensely useful to pursue. With their permission, such self-reports can gain invaluable knowledge about SnehAI users' characteristics, such as age, gender, geographic location, and socioeconomic status. In particular, this user profiling approach can provide empirical evidence about several affordances discussed in this paper, namely, accessibility, inclusivity, glocalizability, and scalability.

Additional theoretical models of technology acceptance [69], uses, and gratifications [70] may further deepen our understanding of key factors that help individuals decide to adopt chatbots such as SnehAI, the specific cognitive, affective, and social needs that are fulfilled, and reasons why certain users would choose to return and continue their conversations with the chatbot repeatedly. These aspects of user motivations and behaviors can be assessed through cross-sectional or longitudinal surveys and triangulated with in-depth interviews. Comparing different groups, such as nonusers, light users, and heavy users, can provide additional insights into psychological and contextual determinants.

Although SnehAI, version 2.0, is a stand-alone app, the avatar is still based on the protagonist of *Main Kuch Bhi Kar Sakti Hoon*, and much of its user content is still curated from 3 seasons of the popular television serial drama—183 episodes, 30 minutes in duration and transmedia extensions on digital platforms. This connection gave SnehAI chatbot a compelling face or interface and connected it to a much larger background story with rich SRH information and effective behavior modeling [56,57]. To better study the social impact of this AI chatbot, a field- and web-based experiment can be conducted to test the differences between SnehAI as a stand-alone SRH intervention, SnehAI as a transmedia edutainment extension, and conventional SRH education without SnehAI among Indian adolescents and young adults that need it the most.

Conclusions

As the first *Hinglish* (Hindi and English) AI chatbot deliberately designed for social and behavior change communication, SnehAI is an innovative, unique, and promising app for engaging vulnerable and hard-to-reach populations groups in the context of SRH education and discussion. It offered a private, nonjudgmental, and safe space for users to talk about otherwise sensitive topics, obtain accurate and trustworthy information, access national services through toll-free helplines, and seek personalized consultations. It also opened up exciting opportunities to leverage the emotional appeal of storytelling through thoughtful yet entertaining content, positive outlook of an avatar, relatable verbal and nonverbal expressions, and friendly tone of voice to effectively engage young people in talking and learning about SRH. The comprehensive checklist of affordances and critical user engagement analytics from this case study are not only a powerful testimonial of SnehAI itself but also a significant representation of the potential and impact of AI technologies on social good.

Acknowledgments

SnehAI (version 1.0) was developed with funding from the United Nations Population Fund and SnehAI (version 2.0) was developed with funding from the Bill and Melinda Gates Foundation through the Centre for Social and Behaviour Change in India. The authors are grateful for the tremendous support from Abhijit Mali, Urvashi Mitra, Tanushree Sengupta, Alok Vajpeyi, Ritesh Laddha, Nikita Serrao, and Mandira Kalra Kalaan at the Population Foundation of India; Kriti Sharma and Fintan Naggle at Artificial Intelligence for Good, United Kingdom; Velotio Technologies Pvt Ltd, India; and Yishin Wu, affiliated with the University at Buffalo, State University of New York, New York, United States.

Authors' Contributions

HW conceived the idea for this study and led to the collaborative process of data processing and manuscript writing. SG, AS, SS, PS, and AP reviewed the manuscript multiple times and provided critical comments and revisions. PM and SS shared seminal project documents and represent the Population Foundation of India, owner of the SnehAI chatbot. PS provided the screenshots in Figure 1. SG created Figures 2-7. PS and AP extracted the analytics data, conducted a preliminary analysis, and validated the results reported in this paper. HW created Multimedia Appendices 1-3. SG assisted with translation from Hindi to English and created the video linked to Multimedia Appendix 4.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Main Kuch Bhi Kar Sakti Hoon Facebook page user insights.
[DOCX File , 248 KB - jmir_v24i1e29969_app1.docx]

Multimedia Appendix 2
The functional affordances of SnehAI chatbot.
[DOCX File , 17 KB - jmir_v24i1e29969_app2.docx]

Multimedia Appendix 3
An illustration of Language Understanding (LUIS) natural language processing system in SnehAI.
[DOCX File , 52 KB - jmir_v24i1e29969_app3.docx]

Multimedia Appendix 4
Video illustration of Language Understanding (LUIS) natural language processing system in SnehAI.
[MP4 File (MP4 Video), 6336 KB - jmir_v24i1e29969_app4.mp4]

References

1. India population (2021). Worldometer. URL: https://www.worldometers.info/world-population/india-population/ [accessed 2021-04-22]
2. Sample Registration System Statistical Report 2018. Office of the Registrar General & Census Commissioner, India. URL: https://censusindia.gov.in/Vital_Statistics/SRS_Report_2018/SRS_Statistical_Report_2018.pdf [accessed 2021-04-22]
3. Jejeebhoy S, Santhya K, Zavier A. Sexual and reproductive health in India. Global Public Health. 2020. URL: https://oxfordre.com/publichealth/view/10.1093/acrefore/9780190632366.001.0001/acrefore-9780190632366-e-225 [accessed 2021-04-22]
4. Muttreja P, Singh S. Family planning in India: the way forward. Indian J Med Res 2018 Dec;148(Suppl):1-9 [FREE Full text] [doi: 10.4103/ijmr.IJMR_2067_17] [Medline: 30964076]
5. Tripathi N. Does family life education influence attitudes towards sexual and reproductive health matters among unmarried young women in India? PLoS One 2021 Jan 25;16(1):e0245883 [FREE Full text] [doi: 10.1371/journal.pone.0245883] [Medline: 33493186]
6. National Family Health Survey (NFHS-4), 2015-2016. Dhsprogram. URL: https://dhsprogram.com/pubs/pdf/fr339/fr339.pdf [accessed 2021-04-22]
7. UDAYA - understanding the lives of adolescents and young adults: Bihar factsheet 2016. Popcouncil. URL: https://www.popcouncil.org/uploads/pdfs/2017PGY_UDAYA-BiharFactsheet.pdf [accessed 2021-04-22]
8. Adolescent health: priorities and opportunities for Rashtriya Kishor Swasthya Karyakram (RKSK) in Uttar Pradesh. Population Council, New Delhi. URL: https://www.popcouncil.org/uploads/pdfs/2017PGY_UDAYA-RKSKPolicyBriefUP.pdf [accessed 2021-04-22]
9. Gadkari RP, Somani G, Nayak CS, Giri AS. A study for sexual health awareness in adolescent population (13-18 years) attending dermatology OPD. Indian J Sex Transm Dis AIDS 2012 Jul;33(2):148-149 [FREE Full text] [doi: 10.4103/0253-7184.102140] [Medline: 23188948]
10. Muzammil K. Helplines for adolescents in India. In: Mahmood SE, editor. Concise Text Book of Adolescent Health. New Delhi, India: IP Innovative Publication Pvt. Ltd; 2019:108-113.
11. Kemp S. Digital 2020: India. Datareportal. 2020. URL: https://datareportal.com/reports/digital-2020-india [accessed 2021-04-22]
12. Social media for youth and civic engagement in India. UNDP. URL: https://www.undp.org/content/dam/india/docs/UNV/Publications/SOCIAL%20MEDIA%20REPORT.pdf [accessed 2021-04-22]
13. Social media usage in India - statistics and facts. Statista. URL: https://www.statista.com/topics/5113/social-media-usage-in-india [accessed 2021-04-22]
14. Følstad A, Araujo T, Papadopoulos S, Law E, Granmo O, Luger E, et al. Chatbot research and design. In: Proceedings of the Third International Workshop, CONVERSATIONS 2019. 2019 Presented at: Third International Workshop, CONVERSATIONS 2019; November 19–20, 2019; Amsterdam, The Netherlands. [doi: 10.1007/978-3-030-39540-7]
15. Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 1966 Jan;9(1):36-45. [doi: 10.1145/365153.365168]
16. Seymour M, Riemer K, Kay J. Actors, avatars and agents: potentials and implications of natural face technology for the creation of realistic visual presence. J Assoc Inf Syst 2018:953-981. [doi: 10.17705/1jais.00515]

17. Stoeckli E, Dremel C, Uebernickel F, Brenner W. How affordances of chatbots cross the chasm between social and traditional enterprise systems. Electron Mark 2019 Aug 08;30(2):369-403. [doi: 10.1007/s12525-019-00359-6]
18. Kocaballi AB, Berkovsky S, Quiroz JC, Laranjo L, Tong HL, Rezazadegan D, et al. The personalization of conversational agents in health care: systematic review. J Med Internet Res 2019 Nov 07;21(11):e15360 [FREE Full text] [doi: 10.2196/15360] [Medline: 31697237]
19. Laranjo L, Dunn A, Tong H, Kocaballi AB, Chen J, Bashir R, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc 2018 Sep 01;25(9):1248-1258 [FREE Full text] [doi: 10.1093/jamia/ocy072] [Medline: 30010941]
20. Milne-Ives M, de Cock C, Lim E, Shehadeh MH, de Pennington N, Mole G, et al. The effectiveness of artificial intelligence conversational agents in health care: systematic review. J Med Internet Res 2020 Oct 22;22(10):e20346 [FREE Full text] [doi: 10.2196/20346] [Medline: 33090118]
21. Car LT, Dhinagaran DA, Kyaw BM, Kowatsch T, Joty S, Theng Y, et al. Conversational agents in health care: scoping review and conceptual analysis. J Med Internet Res 2020 Aug 07;22(8):e17158 [FREE Full text] [doi: 10.2196/17158] [Medline: 32763886]
22. Rahman R, Rahman MR, Tripto N, Ali M, Apn S, Shafriyar R. AdolesecentBot: understanding opportunities for chatbots in combating adolescent sexual and reproductive health problems in Bangladesh. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. 2021 Presented at: CHI '21: CHI Conference on Human Factors in Computing Systems; May 8 - 13, 2021; Yokohama Japan p. 1-15. [doi: 10.1145/3411764.3445694]
23. Arora P. The Next Billion Users – Digital Life Beyond the West. Cambridge, MA: Harvard University Press; 2020.
24. eNGO Challenge Award for digital tools and empowerment. eNGO Challenge. URL: https://engochallenge.org/population-foundation-of-india/ [accessed 2021-06-01]
25. Population Foundation of India. Outreach and communication engagement using AI-powered chatbot. 2019 Presented at: A Development and Pre-Testing Report Submitted to the United Nations Population Fund; 2019; India.
26. Population Foundation of India. SnehAI: an AI-powered companion for young people. 2020 Presented at: A Program Report Submitted to Centre for Social and Behaviour Change; 2020; India.
27. Singhal A, Rogers E. Entertainment-Education: A Communication Strategy for Social Change. Mahwah, New Jersey: Routledge; 2012.
28. Singhal A, Cody M, Rogers E, Sabido M. Entertainment-Education and Social Change: History, Research, and Practice. Mahwah, New Jersey: Erlbaum Associates; 2004.
29. Gibson J. The Theory of Affordances. New Jersey: Lawrence Erlbaum Associates; 1977:67-82.
30. Gibson J. Ecological Approach to Visual Perception. Boston: Houghton Mifflin; 1979.
31. Hutchby I. Technologies, texts and affordances. Sociology 2001 Jul 02;35(2):441-456. [doi: 10.1177/s0038038501000219]
32. Faraj S, Azad B. The materiality of technology: an affordance perspective. In: Leonardi PM, Mardi BA, Kallinikos J, editors. Materiality and Organizing: Social Interaction in a Technological World. Oxford, United Kingdom: Oxford University Press; 2012:237-258.
33. Reed ES. The intention to use a specific affordance: a framework for psychology. In: Development in Context. New Jersey: Lawrence Erlbaum Associates; 1993:45-75.
34. Withagen R, de Poel HJ, Araújo D, Pepping G. Affordances can invite behavior: reconsidering the relationship between affordances and agency. New Ideas Psychol 2012 Aug;30(2):250-258. [doi: 10.1016/j.newideapsych.2011.12.003]
35. Evans SK, Pearce KE, Vitak J, Treem JW. Explicating affordances: a conceptual framework for understanding affordances in communication research. J Comput-Mediat Comm 2016 Dec 26;22(1):35-52. [doi: 10.1111/jcc4.12180]
36. Norman DA. The Psychology of Everyday Things. New York: Basic Books; 1988.
37. Norman D. The Design of Everyday Things: Revised and Expanded Edition. New York: Basic Books; 2013.
38. Conole G, Dyke M. What are the affordances of information and communication technologies? Res Learn Technol 2004 Jun 01;12(2). [doi: 10.3402/rlt.v12i2.11246]
39. Wellman B, Quan-Haase A, Boase J, Chen W, Hampton K, Díaz I, et al. The social affordances of the internet for networked individualism. J Comput Mediat Commun 2003;8(3):JCMC834. [doi: 10.1111/j.1083-6101.2003.tb00216.x]
40. Schrock AR. Communicative affordances of mobile media: portability, availability, locatability, and multimediality. Int J Commun 2015;9:1229-1246 [FREE Full text]
41. boyd D. Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. New York: Routledge; 2010:39-58.
42. Merolli M, Gray K, Martin-Sanchez F, Lopez-Campos G. Patient-reported outcomes and therapeutic affordances of social media: findings from a global online survey of people with chronic pain. J Med Internet Res 2015 Jan 22;17(1):e20 [FREE Full text] [doi: 10.2196/jmir.3915] [Medline: 25616273]
43. Moreno MA, D'Angelo J. Social media intervention design: applying an affordances framework. J Med Internet Res 2019 Mar 26;21(3):e11014 [FREE Full text] [doi: 10.2196/11014] [Medline: 30912754]
44. Strong D, Volkoff O, Johnson S, Pelletier L, Tulu B, Bar-On I, et al. A theory of organization-EHR affordance actualization. J Assoc Inf Syst 2014 Feb;15(2):53-85. [doi: 10.17705/1jais.00353]

45. Treem J, Leonardi P. Social media use in organizations: exploring the affordances of visibility, editability, persistence, and association. SSRN J 2012 [FREE Full text] [doi: 10.2139/ssrn.2129853]
46. Bucher T, Helmond A. The affordances of social media platforms. In: Burgess J, Poell T, Marwick A, editors. The SAGE Handbook of Social Media. Thousand Oaks, CA: SAGE Publications; 2017:233-253.
47. Crutzen R, Peters GY, Portugal SD, Fisser EM, Grolleman JJ. An artificially intelligent chat agent that answers adolescents' questions related to sex, drugs, and alcohol: an exploratory study. J Adolesc Health 2011 May;48(5):514-519. [doi: 10.1016/j.jadohealth.2010.09.002] [Medline: 21501812]
48. Contractor NS, Singhal A, Rogers EM. Metatheoretical perspectives on satellite television and development in India. J Broadcast Electron Media 1988 Mar;32(2):129-148. [doi: 10.1080/08838158809386690]
49. Go E, Sundar SS. Humanizing chatbots: the effects of visual, identity and conversational cues on humanness perceptions. Comput Hum Behav 2019 Aug;97:304-316. [doi: 10.1016/j.chb.2019.01.020]
50. Ischen C, Araujo T, Voorveld H, Noort G, Smit E. Privacy Concerns in Chatbot Interactions. Cham: Springer; 2020:34-48.
51. Liu B, Sundar SS. Should machines express sympathy and empathy? Experiments with a health advice chatbot. Cyberpsychol Behav Soc Netw 2018 Oct;21(10):625-636. [doi: 10.1089/cyber.2018.0110] [Medline: 30334655]
52. Feine J, Gnewuch U, Morana S, Maedche A. Gender Bias in Chatbot Design. Cham: Springer; 2020:79-93.
53. Graber C. Artificial intelligence, affordances and fundamental rights. In: Life and the Law in the Era of Data-Driven Agency. Cheltenham, United Kingdom: Edward Elgar Publishing; 2020:194-213.
54. Lessig L. Code: And Other Laws of Cyberspace Version 2.0. New York: Basic Books; 2006.
55. Wang H, Singhal A. East los high: transmedia edutainment to promote the sexual and reproductive health of young Latina/o Americans. Am J Public Health 2016 Jun;106(6):1002-1010. [doi: 10.2105/ajph.2016.303072]
56. Wang H, Singhal A. Mind the Gap! Confronting the Challenges of Translational Communication Research in Entertainment-Education. London, UK: Palgrave Macmillan; 2021:223-242.
57. Wang H, Singhal A. Theorizing Entertainment-Education: A Complementary Perspective to the Development of Entertainment Theory. Oxford, United Kingdom: Oxford University Press; 2021:819-838.
58. Wang H, Singhal A. Audience-centered discourses in communication and social change: the ‘Voicebook’ of Main Kuch Bhi Kar Sakti Hoon, an entertainment-education initiative in India. J Multicult ral Discours 2018 Jun 21;13(2):176-191. [doi: 10.1080/17447143.2018.1481857]
59. Wang H, Singhal A, Muttreja P, Vajpeyi A, Laddha R, Singh S, et al. The power of narrative persuasion: how an entertainment-education serial drama tackled open defecation and promoted contraceptive use in India. J Develop Commun 2020;31(2):1-13 [FREE Full text]
60. Main Kuch Bhi Kar Sakti Hoon, reel to real - Swachhata Elaan (Video Playlist). YouTube. 2020. URL: https://www.youtube.com/watch?v=ohFI02eJ0wA&list=PLdGZB12BOnbvgj5yExB2QGV7VDaW9IK4L [accessed 2021-04-22]
61. What is language understanding (LUIS). Microsoft. URL: https://docs.microsoft.com/en-us/azure/cognitive-services/luis/what-is-luis [accessed 2021-04-22]
62. Yin RK. Case Study Research, Design and Method 4th ed. Thousand Oaks, CA: Sage Publications; 2008.
63. Crowe S, Cresswell K, Robertson A, Huby G, Avery A, Sheikh A. The case study approach. BMC Med Res Methodol 2011 Jun 27;11(1):100 [FREE Full text] [doi: 10.1186/1471-2288-11-100] [Medline: 21707982]
64. Stake RE. The Art of Case Study Research. Thousand Oaks, CA: SAGE Publications; 2008.
65. Dashbot. URL: https://www.dashbot.io/ [accessed 2021-04-22]
66. Machine-learning features. Microsoft. URL: https://docs.microsoft.com/en-us/azure/cognitive-services/luis/luis-concept-feature [accessed 2021-04-22]
67. Horton D, Wohl R. Mass communication and para-social interaction: observations of intimacy at a distance. Psychiatry 2016 Nov 08;19(3):215-229. [doi: 10.1080/00332747.1956.11023049]
68. Banks J, Bowman ND, Lin JT, Pietschmann D, Wasserman JA. The common player-avatar interaction scale (cPAX): Expansion and cross-language validation. International Journal of Human-Computer Studies 2019 Sep;129:64-73. [doi: 10.1016/j.ijhcs.2019.03.003]
69. Venkatesh V, Thong J, Xu X. Consumer acceptance and use of information technology: extending the unified theory of acceptance and use of technology. MIS Q 2012;36(1):157. [doi: 10.2307/41410412]
70. Ruggiero TE. Uses and gratifications theory in the 21st century. Mass Commun Society 2000 Feb;3(1):3-37. [doi: 10.1207/s15327825mcs0301_02]

Abbreviations

AI: artificial intelligence
GIF: graphics interchange format
LUIS: Language Understanding
NLP: natural language processing
RKSK: Rashtriya Kishor Swasthya Karyakram
SRH: sexual and reproductive health

Edited by R Kukafka, G Eysenbach; submitted 27.04.21; peer-reviewed by A Gesser-Edelsburg, R Zhang; comments to author 22.05.21; revised version received 31.07.21; accepted 21.11.21; published 03.01.22.

Please cite as:
Wang H, Gupta S, Singhal A, Muttreja P, Singh S, Sharma P, Piterova A
An Artificial Intelligence Chatbot for Young People's Sexual and Reproductive Health in India (SnehAI): Instrumental Case Study
J Med Internet Res 2022;24(1):e29969
URL: https://www.jmir.org/2022/1/e29969
doi:10.2196/29969
PMID:34982034

Original Paper

Barriers and Facilitators to Accessing Digital Health Tools Faced by South Asian Canadians in Surrey, British Columbia: Community-Based Participatory Action Exploration Using Photovoice

Antonia Hyman[1], MSc; Elizabeth Stacy[1], MA; Humaira Mohsin[1], PhD; Kaitlin Atkinson[1], MPH; Kurtis Stewart[1], MSc; Helen Novak Lauscher[1], PhD; Kendall Ho[1], MD, FRCPC

Digital Emergency Medicine, Department of Emergency Medicine, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada

Corresponding Author:
Kurtis Stewart, MSc
Digital Emergency Medicine
Department of Emergency Medicine, Faculty of Medicine
University of British Columbia
3rd Floor (Blusson Spinal Cord Centre)
818 West 10th Avenue
Vancouver, BC, V5Z 1M9
Canada
Phone: 1 6048228757
Email: kurtis.stewart@ubc.ca

Abstract

Background: South Asian community members in Canada experience a higher burden of chronic disease than the general population. Digital health innovations provide a significant opportunity to address various health care challenges such as supporting patients in their disease self-management. However, South Asian community members are less likely to use digital tools for their health and face significant barriers in accessing them because of language or cultural factors.

Objective: The aim of this study is to understand the barriers to and facilitators of digital health tool uptake experienced by South Asian community members residing in Canada.

Methods: This study used a qualitative community-based participatory action research approach. Residents from Surrey, British Columbia, Canada, who spoke 1 of 4 South Asian languages (Hindi, Punjabi, Urdu, or Tamil) were invited to participate in focus group discussions. A subsample of the participants were invited to use photovoice methods in greater depth to explore the research topics.

Results: A total of 197 participants consented to the focus group discussions, with 12 (6.1%) participating in the photovoice phase. The findings revealed several key obstacles (older age, lack of education, and poor digital health literacy) and facilitators (social support from family or community members and positive attitudes toward technology) to using digital health tools.

Conclusions: The results support the value of using a community-based participatory action research approach and photovoice methods to engage the South Asian community in Canada to better understand digital health competencies and needs. There were several important implications for policy makers and future research, such as continued engagement of community leaders by health care providers and administrators to learn about attitudes and preferences.

(J Med Internet Res 2022;24(1):e25863) doi:10.2196/25863

KEYWORDS

immigrants; community-based participatory action research; eHealth; delivery of health care; photovoice; South Asian; digital health; mobile phone

Introduction

Background

The South Asian population is one of the largest and fastest-growing ethnic minority groups in Canada, comprising nearly 2 million individuals (5.6% of the general population) in 2016 and increasing annually [1]. South Asian immigrants (ie, people whose ethnic roots originate from the Indian subcontinent, including India, Pakistan, Sri Lanka, Nepal, Bangladesh, Maldives, and Bhutan) experience a higher burden of chronic diseases than the general population (eg, cardiovascular disease and type 2 diabetes) [2,3], tend to have more risk factors for chronic diseases [4,5], and face many challenges accessing tools to prevent chronic disease and manage their health conditions [6,7]. Consequently, there is a pressing need to address barriers to health care access for South Asian community members [8] and reduce health inequities [9].

Digital Health Tools

The prevalence and widespread availability of digital health (eHealth) innovations provide a significant opportunity for addressing various health care challenges [10]. A systematic review of literature on *eHealth* attempted to understand the various ways it can be defined and found 51 different definitions [11]. The authors of the review concluded that developing another definition in an attempt to improve or summarize previous definitions did not make sense, given that specific wording will have a place in different contexts, and therefore did not merit delineating further. For the purposes of this research, digital health can be broadly defined as the transfer and delivery of health care and health resources and services through information and communication technologies [12]. Digital health *tools* specifically are considered the physical tools required to access digital health technology. For example, the use of electronic health records, making medical appointments or filling prescriptions on the web, telehealth for rural patients, or using a smartphone app to track one's health status. It is estimated that if all Canadians had access to digital health tools, it would decrease 47 million in-person visits to health care providers and 18.8 million hours in absenteeism from work, representing an annual gain of CAD $400 million (US $312 million) in gross domestic product, and provide 51 million additional hours to spend on leisure activities [13]. Research on digital health and the impact on quality of care tends to measure quality of care from the health care professional or management perspective. For instance, a large retrospective study compared quality of care at sites using electronic health records versus those using only paper-based records for patients with chronic health conditions [14]. The study found an improvement in quality of care and outcomes according to regional standards of care. The successful incorporation of digital health tools into one's health care activities relies on appropriate levels of digital health literacy [15], which is based on the intersection of six foundational literacies: traditional literacy and numeracy as well as computer, health, information, media, and science literacies.

Despite the benefits, there can be significant barriers to the uptake of digital health tools by individuals, such as sociotechnical challenges [16] or misalignment between design and user needs [17]. In particular, immigrants and older adults can face specific challenges concerning access and uptake of digital health tools [12]. In a sample of Punjabi-speaking South Asian individuals in Canada, older age, female gender, lack of language proficiency, and lower socioeconomic status were all associated with less use of technology for health self-management [12]. A large survey conducted in Alberta, Canada, gives a picture of technology and digital health tool use among South Asian Canadian adults. It found that 74.5% of the participants reported using the internet, with 47.8% using it for health information–seeking purposes; 74.9% reported using smartphones, and of these respondents, 30.7% had apps related to health and fitness [18]. The survey also found that older age and lower educational attainment predicted lower use of the internet and smartphones, among other factors. In addition, the survey found that preferring languages other than English predicted lower likelihood of using different forms of eHealth. Although digital health innovations exist to support prevention and management of chronic diseases, these tools are often not culturally tailored or used by South Asian populations in Canada [12,19]. This gap could lead to disparities in health knowledge and services [20]. As such, implementing digital health–related innovations among South Asian Canadians requires understanding and addressing barriers and facilitators that they may face when using digital health tools.

Community-Based Participatory Action Research and Photovoice

Community-based participatory action research (CBPAR) methods [21,22] have several advantages compared with conventional researcher-driven methods. By co-designing and co-leading the advancement of research between community members and researchers, CBPAR can empower and build capacity in the community and acknowledges that community members are experts in their own right concerning issues affecting them [23]. Thus, CBPAR can foster understanding of research from the participants' point of view and capture important insights that may be missed by investigators external to the community.

Photovoice is a process that uses photography and discussion to bring attention to community issues and aims to empower the community of interest [24,25]. This methodology invites participants to take photographs in their daily lives that capture concepts important to the research question [26], and they write the meaning of the photograph in their own words. Once the set of photographs is captured, a group discussion follows to share the meaning behind the photographs. A review of studies using photovoice illustrates that this approach can engage communities in meaningful research to promote positive change, even in hard-to-reach populations [23]. Photovoice can overcome potential barriers related to language and literacy through the emphasis on photographs [26], and therefore it was identified as a valuable method for this study where such barriers may exist among participants.

This Study: Interactive Health Education Action for Life (iHEAL)

We conducted this 2-phase CBPAR study to better understand the barriers to and facilitators of digital health uptake among the South Asian Canadian community. This study engaged the South Asian community in Surrey, a municipality in Metro Vancouver in western Canada with a large South Asian population. At the time of data collection, Surrey had a total population of 517,885, with a growth rate of 11% from 2011 to 2016 (compared with 7% for Metro Vancouver) [27]. Immigrants make up 43% of the population, with 41% originating from India, and South Asian individuals make up one-third of the population [27]. The regional health authority has identified that South Asians have poorer reported health than the general population, with type 2 diabetes and cardiovascular disease 2-3 times higher and 46% of the older adults having ≥2 chronic diseases [5].

Research Aims

This iHEAL study has 3 primary objectives. First, it aims to gain a deeper understanding of the barriers and facilitators experienced by Surrey's South Asian community in using digital health tools for self-management and prevention of chronic disease. Second, it aims to provide an opportunity to engage the community and increase awareness and capacity with research and technology. Third, it aims to illuminate key policy areas related to the development and uptake of digital health tools in this population.

Ethical Approval

All procedures performed with participants in this study were in accordance with the ethical standards of the Behavioural Research Ethics Board of The University of British Columbia, which provided ethics approval for conducting the study (H14-02308), and with the 1964 Declaration of Helsinki and its later amendments. Informed consent was obtained from all individual participants included in this study.

Methods

Study Design

This iHEAL study comprised two phases: (1) focus group discussions and (2) photovoice. Given the diversity within Surrey's South Asian community and the different languages spoken, the methods purposively targeted communities speaking Punjabi, Hindi, Urdu, and Tamil to represent a plurality of views [27]. Guided by community-based participatory action methodology [21,22], the team established an advisory committee to expand the engagement of community organizational leaders as well as to recruit 7 peer community researchers (PCRs) from the South Asian community to provide culturally appropriate research support for the project and build research capacity in the community itself.

The research team sought to ensure that the initiative was truly co-designed and co-led with the community. The research group had a long-standing relationship in Surrey's South Asian community, with multiple researchers on the team being from the community itself. These researchers came with a wide variety of perspectives and expertise. The initial phases of the initiative were informed by two groups: (1) a broad group of stakeholders with various interests, agendas, and mandates and (2) leaders from various community-based organizations. Both groups aimed to collaboratively identify key research questions, target populations, methods of engagement, and other aspects core to the research initiative. These sessions were facilitated in the language of the community by members of the community. The research tools developed with partners were translated into the target languages, after which a small cadre of partners reviewed them to ensure that the questions made sense and resonated. For instance, a question regarding any differential impact of gender roles and experiences of digital health tools was included because this was identified as an important area to explore by the community. The community partners conducted focus groups, facilitated photovoice sessions, analyzed data, and disseminated findings.

Participants and Sampling Method

Recruitment was facilitated by a community liaison, community leaders, PCRs, and The University of British Columbia research team. Language-appropriate posters and recruitment booths were hosted by the team at community centers, schools, and places of worship. Eligible participants were adults (aged ≥18 years) of self-identified South Asian background who were able to converse in 1 of the 4 target languages (Punjabi, Hindi, Urdu, or Tamil) or English. Taking into account the population who face the highest barriers and challenges to integrating digital health tools into their self-management regime [12,18], recruitment primarily targeted older adults. However, the inclusion criterion was kept at age ≥18 years to accommodate other participants interested in joining the study or younger family members of participants interested in joining the study. Notably, older adults are often supported in their technology needs by other family members in South Asian households. Interested individuals were provided with further information about the study and given time to consider and voice any questions or concerns, after which they could give informed consent if they wished to participate.

Procedures and Data Collection

Phase 1

Focus group discussions were semistructured, with questions designed to explore barriers to and facilitators of digital health tool use. A trained focus group moderator led each discussion using a template with 5 primary questions and follow-up questions where clarification was necessary (Textbox 1). Each focus group discussion was 60 to 90 minutes in duration and audio recorded with participants' consent. The 7 PCRs guided the focus group discussions and provided language support as needed.

Textbox 1. Primary questions for the phase 1 focus group discussions.

Primary questions and possible follow-up questions and prompts

- What is one important health related priority you have for you and your family?
  - What are some of the obstacles you come across in staying healthy? Think of any and all aspects of health such as physical health and mental health.
- How do you most commonly access the health information you need?
  - Why do you choose to use those resources in particular?
  - Whom do you trust most when deciding what information will help you manage your health?
- Does technology play a role helping you manage your health?
  - What do you use technology to help with specifically?
  - What stops you from using technology?
  - What makes some technologies easier or harder to use than others?
- What has your experience been with health information on the web?
  - Do men and women experience the same challenges to using technology to obtain health information?
  - What do people see as concerns and barriers (for example, is the material available in your own language?)
- How is the current health system supporting or interfering with your ability to manage your health?
  - How can the health system support you to use digital health technology to better manage your health?

Phase 2

We sought to recruit 12 participants from phase 1 into the subsequent photovoice phase, given that this sample size is the *gold standard* for the photovoice process and ensures a proportional selection of participants from each language community [21,22]. Orientation sessions were held to review the study purpose, time commitment, and procedures, as well as to answer questions. The sessions explained how to use the cameras provided and gave participants the opportunity to practice taking, accessing, and deleting photographs. Each participant was assigned a PCR to provide support for the duration of this phase. Participants were given 14 days to take 10-15 photographs that reflected their experiences of using technology to improve their own health and the health of their community. Three discussion questions to guide the photography activity were developed by the researchers and PCRs to ensure cultural appropriateness:

1. What comes to mind when you think about using technology to learn more about health or manage it?
2. What difficulties do you face when using technology to learn more about health or manage it?
3. What helps you use technology to learn more about health or manage it?

The photographs were printed and used in the focus group discussions with the consent of participants.

Data Analysis

The focus group discussions from both studies were professionally translated and transcribed verbatim. A constant-comparison method [28-30] was used to analyze the focus group transcripts using NVivo software (version 10; QSR International). Anonymized data were uploaded onto NVivo, and the data from all focus groups were compiled and, for each phase, analyzed separately in 3 stages. Stage 1 involved chunking the data into smaller units and coding them into a few words or short paraphrased text according to the essence of each individual quote. Stage 2 involved grouping these codes into wider categories, and stage 3 involved regrouping or merging related text to create overarching themes. This approach is useful when analyzing data across multiple focus groups, enabling the comparison of data across groups to see if similar themes emerge and check for saturation of data to indicate the salience of themes across the entire sample. The data were independently analyzed by 2 researchers (KA and HM), and themes were checked and cross-referenced by a third researcher (AH) to determine reliability and validity. The final themes were selected based on consensus and the quantity and quality of supporting data. A researcher involved in the data analysis was from the South Asian community and provided additional insight to contextualize any themes that were culturally specific. The themes were presented back to participants at a presentation as part of an event to showcase the photographs taken in phase 2 (with consent), and feedback was obtained from participants indicating that the themes captured the essence of the discussions and did not miss key information.

Results

Overview

Phase 1 data were collected from March 2015 to June 2016. A total of 197 individuals residing in Surrey participated: 81 (41.1%) Punjabi-, 67 (34%) Hindi-, 35 (17.8%) Urdu-, and 14 (7.1%) Tamil-speaking participants. In all, 26 focus groups were conducted (mean 7.6, SD 1.8 participants per group), with 13,

7, 4, and 2 focus groups conducted for the Punjabi-, Hindi-, Urdu-, and Tamil-speaking communities, respectively. During the initial focus groups, the moderators observed that men spoke much more than women; therefore, subsequent focus groups were separated by self-reported gender to promote a balance of perspectives. For phase 2, of the 197 participants, 12 (6.1%) completed the photovoice component, and 2 focus groups were conducted in July 2016 to share photographs with the group and discuss their meaning.

Demographic Characteristics

For phase 1, a total of 197 people participated with consent; however, demographic data were completed by only 130 (66%) participants (Table 1). The participants had a mean age of 65.4 years (SD 12.1 years, range 31-90 years), with a large proportion of older adults. Almost all participants were born outside of Canada, with most of them residing in the country for at least 10 years. Most of the participants (111/130, 85.4%) indicated that their physician was their primary source of health information, and approximately half reported that the internet, family and friends, and television and radio were secondary sources.

Table 1. Phase 1 participants' demographics by language group (N=130)[a].

Demographics	Punjabi (n=45)	Hindi (n=47)	Urdu (n=23)	Tamil (n=13)	Total (N=130)
Age (years), n (%)					
<50	6 (13)	0 (0)	2 (9)	9 (75)	19 (15)
51-60	2 (4)	4 (9)	2 (9)	0 (0)	8 (6)
61-70	14 (31)	25 (53)	10 (46)	3 (25)	52 (41)
>70	23 (51)	18 (38)	8 (36)	0 (0)	49 (38)
Missing	0	0	1	1	2
Gender, n (%)					
Female	27 (60)	23 (55)	7 (32)	6 (46)	64 (52)
Missing	0	5	1	0	6
Education, n (%)					
Secondary or below	15 (34)	11 (25)	10 (50)	1 (8)	37 (30)
Diploma	4 (9)	10 (23)	4 (20)	1 (8)	19 (16)
Undergraduate	17 (39)	12 (27)	3 (15)	4 (33)	36 (30)
Postgraduate	8 (18)	11 (25)	3 (15)	6 (50)	30 (25)
Prefer not to say	1	3	3	1	8
Household income per year (CAD $; US $), n (%)					
<40,000 (31,241.10)	12 (27)	21 (57)	13 (87)	3 (38)	49 (62)
40,000-60,000 (31,241.10-46,861.60)	4 (22)	9 (24)	1 (7)	0 (0)	15 (19)
>60,000 (46,861.60)	2 (11)	7 (19)	1 (7)	5 (63)	15 (19)
Missing or prefer not to say	27	10	8	5	48
Place of birth, n (%)					
Born outside of Canada	43 (98)	46 (100)	23 (100)	13 (100)	127 (99)
Born in Canada	1 (2)	0 (0)	0 (0)	0 (0)	2 (2)
Missing	1	1	0	0	2
Resident in Canada (years), n (%)					
<5	3 (7)	0 (0)	2 (9)	4 (31)	9 (7)
5-10	5 (11)	5 (11)	3 (13)	3 (23)	16 (13)
11-20	8 (18)	9 (20)	6 (26)	3 (23)	28 (22)
>20	28 (64)	31 (69)	12 (52)	3 (23)	74 (58)
Participated in phase 2, n (%)	5 (11)	4 (9)	1 (4)	2 (15)	12 (9)

[a]Percentages were calculated after removing missing and prefer not to say responses from the denominator. In all, 2 participants included in the Total column were missing their language group.

Themes

The qualitative analysis of the focus groups revealed various themes that were consistently identified in both phases. Reported here are first the themes common to both phases, followed by the themes unique to each phase. Finally, additional themes that were specific to phase 2 regarding the photovoice component are reported. Table 2 presents a summary of the core themes identified.

Table 2. Summary of barriers and facilitators to accessing digital health tools.

Summary	Facilitators	Barriers

Common themes across both phase 1 and phase 2	• Social support (*eg, from family members to use digital health tools*)[a]
	• Positive attitude toward using digital health
	  • *Subthemes for positive attitudes*:
	  • Convenience (*eg, saves travelling time*)
	  • Enhances awareness of health (*eg, monitoring activity levels through wearables*)
	  • Reduces need for hospitalization (*eg, sleep apnea machines can be used at home*)
	• Provides opportunities
	  • To learn about health (*eg, reading on the internet about how to manage health conditions*)
	  • For social connection (*eg, using Skype to connect with friends and family*)
	  • For physical activity and mental stimulation (*eg, technology use for games and activities*)
		• Language (*eg, English as a second language*)
		• Literacy levels (*eg, medical terminology*)
		• Computer and technology literacy (*not having learned to use computers*)
		• Recognition of a negative impact on health (*associating time spent on the internet as being unhealthy*)
		• Time constraints (*lack of time to learn to use digital health tools*)

Unique to phase 1 (focus groups only)	• Familiar platforms and places (*for learning and knowledge sharing*)	• Age-related self-efficacy (*lack of confidence in perceived ability to learn because of age*) • Gender roles (*eg, women reporting less digital health tool use because of caring and familial responsibilities*) • Limited trust in digital health sources (*many only trusting physicians' word*) • Cultural norms of communication (*preference for in-person interaction*)
Unique to phase 2 (photovoice)	• Computer classes (*willingness to learn starting with computer classes*) • Usability of technology (*accessibility and user-friendly technology encouraged digital health tool use*) • Social sharing (*of health information has potential to promote use of digital health tools*)	• Financial (*many reported financial constraints to accessing tools*) • General self-efficacy (*general lack of confidence reported in ability to use tools*) • Lack of motivation (*some reported a lack of desire to put effort into trying to use tools*) • Lack of awareness (*many reported not knowing much about the tools available and how to access them*)

[a]Italicized results are further described in the subsections below.

Common Facilitators Across Both Phases

Social Support

Support from family, especially younger generations, was frequently cited as a facilitator to using digital health tools: " Online...I [get] help from my kids. I can't use [a] laptop, but [my] kids always help me whenever I need to know something." There was agreement across many participants that family is a great source of support in using technology: "[I] think most grandchildren, sons and daughters all provide us help in that case."

Positive Attitude Toward Using Digital Health

In phase 1, participants identified more than 20 advantages for using digital health tools, for example:

> *It is very useful for health. It is good for general knowledge. You can see the whole world through [the] internet. It is a need of today's life.*
>
> *What I have noticed is after coming here from India, people are becoming more active physically and mentally. Only technology has made them more active.*

Despite the challenges identified, participants largely agreed on the importance of technology, especially for those who have limited mobility: "To learn [the] internet is very necessary for those seniors who have difficulty in moving around."

In phase 2, participants also identified numerous advantages to using digital health tools:

- Convenience: several participants photographed digital health tools such as pedometer devices, blood pressure machines, and blood glucose monitors (Figure 1) and talked

about the convenience of health technology: "Now I can check my blood pressure at home whenever I have a need to do it...."

- Enhances awareness of health: discussion around the photograph of a Fitbit device sparked conversation about the power of technology to enhance awareness of health:

> *Awareness also [about] how much [exercise] you have done, how much you should do: it helps. If the steps are less [than they should be] you can walk more and if it's more then you don't exert anymore.*

- Provides opportunities to learn about health: many participants referenced ways in which technology can provide opportunities to learn about health, for example:

> *If there is any health issues or disease, we can go to the internet and see why it has happened, the precaution we need to take and diet we need to follow. We can know everything from the internet.*

Some participants spoke about the advantage of being able to read health information on the web in different languages:

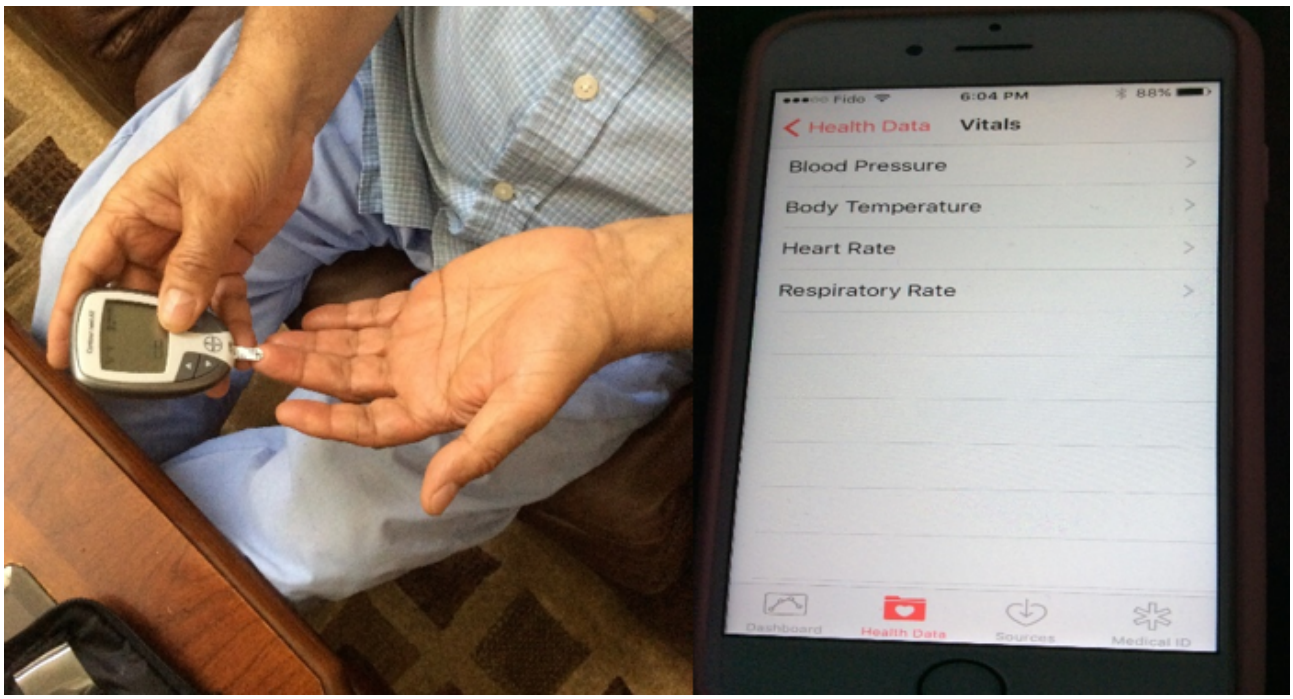
> *If you bring up an article it will ask you...if you want to translate. When you say yes, it will give you options to choose the language.*

- Reduces need for hospitalization: discussion around a photograph of a sleep apnea device revealed that several participants felt that technology plays an important role in reducing the need for hospitalization: "In the hospital somebody [with] more serious [illnesses] can get the bed and this patient got the facility at home."
- Provides opportunities for social connection: another theme that arose was that technology can provide opportunities to connect with others:

> *Skype is a wonderful thing for lonely older people...when you want to talk to your relatives you can do that...you can see the person and have great satisfaction.*

- Creates opportunities for physical activity and mental stimulation: another advantage referenced by several participants was the idea that technology can provide opportunities to be active both physically and mentally. For a participant, keeping mentally active played an important role in managing chronic pain: "It keeps me busy and keeps my mind off the pains I have."

Figure 1. Photograph taken by a participant, with an example of a quote from a participant discussing this photograph in the focus group: "Now I can check my blood pressure at home whenever I have a need to do it...making it easy for us to check blood sugar, blood pressure, heart rate, etc."



Common Barriers Across Both Phases

Language

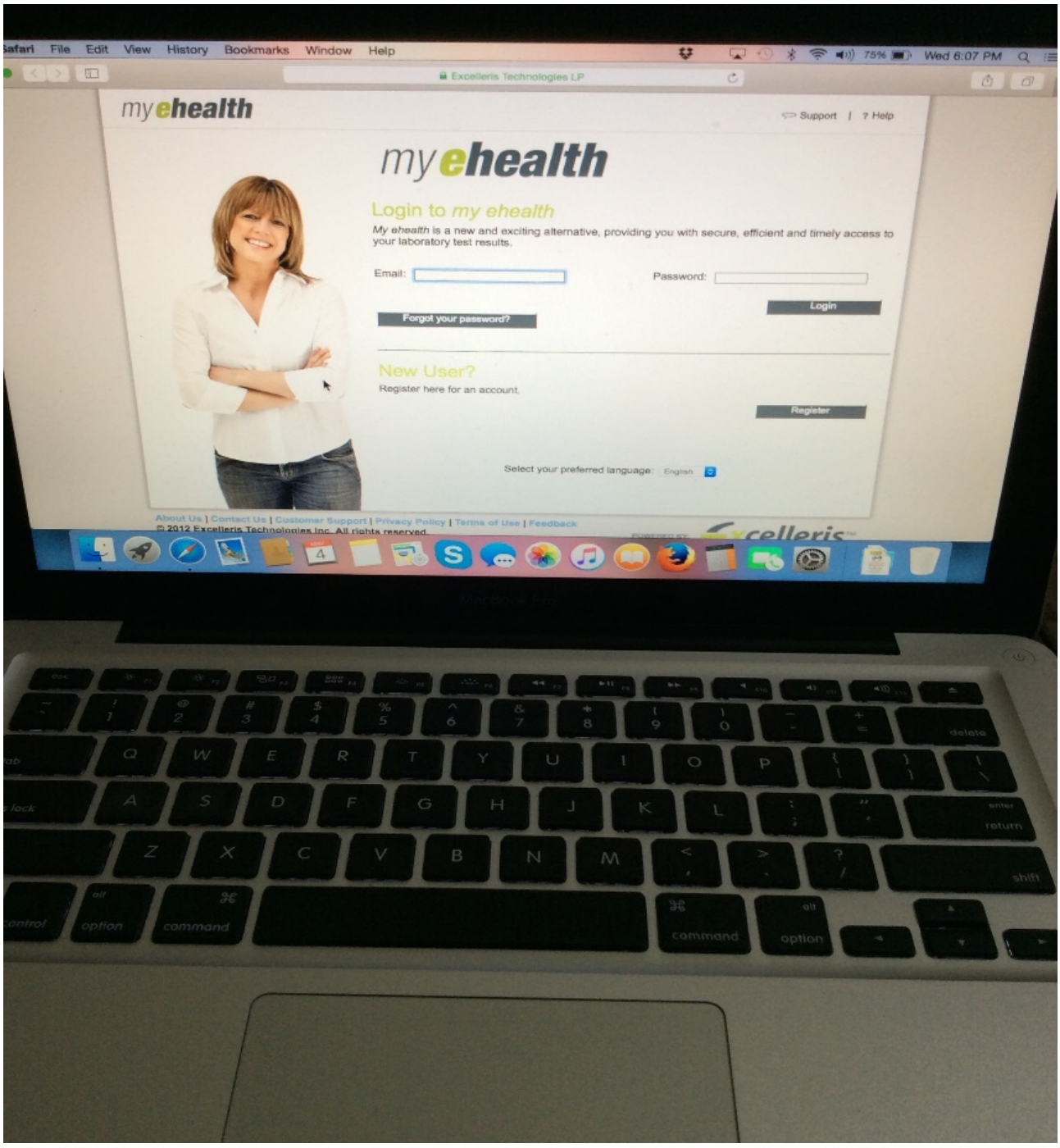
English was a second language for most participants. Even when language-specific resources were provided, participants expressed that the content was often either too academic or not properly translated, for example:

> *It's not simple Punjabi. Complicated words are used in the translations.*

Discussion around a photograph of a computer (Figure 2) also revealed these challenges:

> *You go on [the] internet and everything is in English, whereas it [is] not in their mother tongue. It is a challenge for many.*

Figure 2. Photograph taken by a participant, with an example of a quote from a participant discussing this photograph in the focus group: “Though this picture does not show language, you go on internet and everything is in English, whereas it [is] not in their mother tongue. It is a challenge for many.”



Literacy Levels

Many participants reported that they had difficulty reading in any language and therefore had a preference for radio or television:

> *If you tell me go to [the] internet and access the information but I am not educated, then how will I read that information? You send me literature; I can’t read it so it will go in the garbage. So, the best source is radio and TV.*

Computer and Technology Literacy

Challenges in using technology and navigating web-based information were frequently cited: “I’m an illiterate when it comes to computers.” Some participants expressed an interest in learning to use computers, for example:

> *We don’t know how to use it but nobody even bothered, nobody tried to teach us, neither [did] we put [in the] effort to learn, but if somebody will teach us, we can learn.*

Recognition of Negative Impact on Health

Some participants expressed awareness of the harmful side to excessive technology use: "There are many people who spend too much time on [the] internet and they don't go outside for [a] walk...." Several participants in the photovoice phase noted that excessive technology use can lead to unhealthy behaviors, for example: "If we sit at the computer, three or four hours will go by easily, then later you realize that we have been focusing on only one thing and you have ignored many other things."

Time Constraints

Time was also raised as a barrier to learning to use technology, especially for health: "Actually life is so busy that generally people are not available for sessions or for different aspects of technology especially in the health area..."

Facilitators Unique to Phase 1

Familiar platforms and places—some participants discussed the importance of using familiar gathering places to increase the uptake of digital health tools:

> *To learn about health, in terms of motivating us [South Asian seniors], it has to be easy for them...If they go to community centers, gurdwaras, and other places, that would be the best places to teach the community.*

As part of this conversation, some participants indicated that radio and television are ideal sources for information sharing in the community: "More information should be given through radio and TV."

Barriers Unique to Phase 1

Age-Related Self-efficacy

Concerns of being unable to learn new concepts or having poor memory were reported, which participants often attributed to age:

> *It's of no use now...what [can I] learn at this age?*

Some participants indicated that there is fear of learning for older community members:

> *For all the seniors, it is a phobia to use [the] internet or computer[s]. They think that they are 60+ and they cannot* [learn anything new]

Gender Roles

It was reported that because of women's familial duties, they experienced more limited access to, and time for, seeking health information using digital technologies. Women reported using digital health tools less than men, often attributed to their workload and gender roles in the family: "[Women] have lots of household work as well as work outside [the house], too." However, some participants also indicated that these gender differences in technology use are diminishing:

> *Most of the men use cell phones so they might be able to access [the] internet more than women. But it's getting equal nowadays.*

Limited Trust in Digital Health Sources

Physicians' advice was considered by many participants to be more reliable than the internet or television and radio programs: "Because [my] doctor knows my diagnosis and can tell me the right things." Some participants suggested that they felt it was important to corroborate health information they find on the web with their physician: "So whenever you get any kind of information, you must consult the doctor, before you go for it." Many indicated that they felt that their physician was their primary source of health information: "[My] doctor is the ultimate source."

Cultural Norms of Communication

Face-to-face interaction was preferred over internet-based communication among participants:

> *When email was newly introduced, my next door neighbor sent me an email about something. I felt like he slapped my face. I went to him and asked, "Can't you talk to me about it? Why did you send me [that] in writing?"*

Facilitators Unique to Phase 2

Computer Classes

Many participants agreed that computer classes in the community would enable the use of health technologies: "We need some computer classes because we don't want to depend on others." Participants expressed that an important part of taking the classes is to spark people's curiosity and focus on the interests of the community:

> *You have to start from somewhere. First of all they should start with what they are interested in.*

One-to-one teaching was deemed more effective than group classes by participants, particularly for older adults in the community.

Usability of Technology

Several participants discussed the importance of technology being accessible and easy to use. A participant explained how this made a difference to his uptake of technology:

> *I didn't know what the computer [was]. At that time, you didn't have the easier ways. You didn't have pointer. You couldn't use [a mouse], you had to use certain keys to get into things...So I found it very difficult to learn at that time. Later on they found the easier ways.*"

Furthermore, participants discussed ways in which media can be an accessible digital tool for improving health literacy in the South Asian community (Figure 3). Participants stressed that radio could be a particularly useful tool, especially for older adults:

> *Radio is almost in every senior's home. If not at home then it is accessible on the phones these days.*

Figure 3. Photograph taken by a participant, with an example of a quote from a participant discussing this photograph in the focus group: "There are so many health channels on TV...there is exercise, diet, using equipment, and explain about technology..."

Social Sharing

Another key facilitator to using technology for health was the cultural norm of social sharing among the South Asian community: "When somebody [hears] something [they] would tell another ten people."

Barriers Unique to Phase 2

Financial

Affordability was stated as a clear limiting factor for many people in the community to using technology for health: "If we want to get training ourselves, we don't have money to spend, plus we also need time."

General Self-efficacy

Another barrier discussed was the lack of confidence to use technology:

> *I think it's a fear of confusion. Fear has to be overcome...removed. If you sit at the computer and don't know what to do or how to start, then how can you use it? Then it is very intimidating.*

Lack of Motivation

Motivation was also discussed as an obstacle, largely because of the effort required to learn:

> *They couldn't be bothered. I know the advantages...they think there are some hassles, there is too much trouble to learn.*

Lack of Awareness

Lack of awareness concerning which digital health tools are available. A participant stated as follows:

> *I think the greatest barrier is awareness. If people are not aware of what technology is there to manage their health, they will not use it.*

Experiences of Photovoice

Meaningful and Enjoyable Experience

Participants spoke about the photovoice phase itself as being a meaningful and enjoyable learning experience:

> *It was a good time. We had a chance to see many new things.*

A total of 3 key themes emerged from the focus groups to capture the participants' experiences.

Learning Opportunity

For many participants, using photovoice promoted learning about technology and health in general:

> *There was a lot of exploration...what options are available...how technology is helping in health.*

More specifically, participants' awareness of how technology can be used to promote health seemed to grow through the photovoice process:

> *I really liked it...With new technology, we come to know about so many things. We can come to know how to use technology and how it can help us.*

Personal Growth

Many participants highlighted the novelty of the experience that led to self-reflection and personal growth:

> *The significant moment was when I was taking pictures and started thinking why I was taking the photographs. It opened a field for me. What was happening and why was it happening with me then. It was very enlightening for me.*

Another participant stated as follows: "Actually we gained confidence with it."

Encourages Healthy Behaviors

Among other benefits referenced, some felt it engaged them in a healthy mental activity:

We kept busy. We had [a] kind of brain exercise.

In addition, it encouraged physical activity among some participants: "We get a chance to go out and walk around."

Discussion

Thematic Findings

This multiphase CBPAR study with the South Asian community in Surrey, British Columbia, Canada, investigated facilitators of and barriers to the uptake of digital health tools to prevent and manage chronic health conditions. Some of the barriers identified reflect recent research that immigrants, older adults, and older ethnic minority individuals are less likely to use digital health tools for their health care activities [20]. In addition, participants stated that age, gender, income, and education presented obstacles to their use of digital health tools, which closely echoes the findings from a study of Punjabi-speaking individuals in Metro Vancouver [12]. Perceived gender roles were cited as reasons why *women* may use digital health tools less than *men* because of familial and household duties; however, digital technology can save time and therefore can be a solution for those who have less time or more caring responsibilities. Low English and computer and technology literacies were salient themes and reinforce the need for digital health innovations to be linguistically appropriate and accessible to ethnic minority populations [12,19]. Interestingly, many individuals reported high educational levels (66/122, 54.1%) had an undergraduate or postgraduate degree); even so, a theme emerged around difficulty reading in any language (even first language). A possible explanation for this is that participants were referring more specifically to difficulty understanding medical language, which can be challenging for many people from different cultural and educational backgrounds, particularly considering that those who reported education up to undergraduate and postgraduate level were educated in their country of origin where English language proficiency is limited and often restricted to their professional areas. In addition, we are missing data for background education for 34% (67/197) of the participants; therefore, it could be that this theme represented those with lower levels of education.

A first step in promoting the use of digital health tools in the South Asian community may be to address levels of health and computer literacies. By improving digital health literacy, participants indicated that they would be more likely to start using digital health tools to manage their own health. This important point is intertwined with other themes elucidated, namely that promoting digital health literacy will also help participants to address barriers around their self-efficacy, motivation, and awareness. Community-based promotion efforts could incorporate elements of social sharing, social support, and cultural norms to improve the likelihood of success. Furthermore, there seemed to be some disconnect between the research team's and the participants' perspectives on the technology that constituted *digital health tools*. The participants took a broader view and discussed technology that affected their health, such as the use of videoconferencing for social connection. Thus, it may be necessary to establish a shared definition of *digital health tools* for research and promotion with South Asian communities as well as to focus research on specific digital health tools to be able to take more concrete steps toward improving accessibility of specific tools.

These findings highlight other barriers to the use of digital health tools, namely cultural norms, lack of trust, and time and financial constraints (ie, social determinants of health [SDoH]) [31]. These are important factors to consider and may explain some of the gaps in digital health access for South Asians in Canada. Interestingly, given the number of barriers identified, the participants also seemed to have a generally positive attitude toward digital health tools and were interested in learning more about them and their potential benefits. This reinforces the need for accessible, culturally sensitive digital health tools that are not prohibitively costly in terms of money or time investment. The ways in which tools may be considered culturally sensitive must be guided by the users themselves and not researchers or technology developers. However, this study indicates some ways in which this might be achieved, such as options for first language and perhaps adopting elements of social sharing that do not rely on both or all users accessing technology directly by incorporating a *buddy system*. This may mean that friends and family can benefit from the technology without necessarily having to navigate it directly but through the support of those who feel more confident. Although not specifically discussed in this study, cultural sensitivity must also consider language that may not be appropriate, for instance, language that is more representative of Westernized culture.

The Value of a CBPAR Approach and Photovoice

Our findings support the validity and significance of using CBPAR methods, including photovoice, to develop and explore research questions with ethnocultural populations, while also building knowledge and skills in these communities. Although previous literature has highlighted the risk of CBPAR methods leading to tension between community members and researchers or the possible loss of research objectivity, we concur with experts that CBPAR is a strong approach for promoting health equity, especially in marginalized communities [32,33]. The photovoice participants reported that they increased their awareness of, and confidence in, using digital health tools, which was a success of the capacity-building component of the research initiative. Participants were able to learn about different aspects of health and digital tools through exploration; they had the opportunity to connect and learn with others, while being supported by research team members. The project built research capacity in the community by engaging and mentoring community leaders and PCRs to facilitate the research process, ensuring that all elements were language-appropriate, and making the study meaningful for the community. Although both the focus group and photovoice methods identify common themes, each method further contributed unique insights that provide a more holistic picture of the barriers to and facilitators of the uptake of digital health tools. Although we set out to explore the barriers to and facilitators of digital health tool use, this paper also highlights the challenges of conducting CBPAR in a way that minimizes the potential for bias.

Limitations

Although a large sample was recruited from Surrey's South Asian community, there may be limited generalizability to the wider community of South Asian individuals residing in Canada. For example, South Asian Canadians living in more rural settings or serviced by different regional health care systems may experience other barriers to accessing digital health literacy. Most participants spoke Punjabi or Hindi, the 2 dominant South Asian languages spoken in Surrey [27]. Further work could elaborate on any issues specific to the Urdu- and Tamil-speaking communities (or other minority language groups) as well as to understand how facilitators and barriers may differ among cultural groups or by other demographics. In addition, demographic data beyond language group were missing for a subset of participants, thereby limiting our ability to speculate on the link between the themes and characteristics of the whole sample. The photovoice phase met the gold standard for sample size by involving 12 individuals [23,24]; however, multiple rounds of photovoice with more participants may have gleaned additional information and themes. In addition, we are not able to report on the demographic characteristics of the photovoice participants, thus potentially overlooking deeper gender-based insights. Furthermore, a deeper form of analysis such as discourse analysis could have been valuable to explore any influences that could come from the agenda of community leaders potentially biasing participant responses.

Implications for Policy and Future Work

Considering the increasing prevalence of digital health tools within the current health care landscape, these findings highlight considerations for the use of these tools among South Asian Canadians in terms of accessing health information and managing health conditions and preventing chronic disease. On the basis of this study's findings and past work, there seem to be several policy and health care recommendations, which are organized here according to level [31]:

- Microlevel (patients and caregivers)
  - Continue to assess the current and future health needs of ethnic minority groups accessing health care, including digital health literacy and SDoH.
  - Involve patients and caregivers throughout research and health initiatives to ensure that these efforts are meaningful and culturally appropriate.
- Mesolevel (community)
  - Evaluate community attitudes toward digital health tools to inform optimal approaches in implementing new digital health initiatives.
  - Integrate CBPAR approaches and photovoice as methods of inquiry to ensure that diverse and holistic perspectives are collected [24,26].
  - Incorporate cultural norms and preferences into initiatives to develop and promote digital health literacy to maximize their appeal and accessibility.
  - Develop versions of digital health tools that are not prohibitively costly in terms of money or time investment.
- Macrolevel (education and health system)
  - Engage community leaders, health care providers and administrators, and technology developers to better understand the needs of groups with varying SDoH.
  - Raise awareness among health care providers regarding facilitators and barriers around digital health tool uptake, with the objective of improving providers' communication and prescription of digital tools to patients.
  - Provide opportunities for health professionals and trainees to learn from multicultural patient and community populations to develop greater understanding of potential barriers and cultural considerations.

Conclusions

In the transformation of health systems to introduce digital technology innovations, it is necessary to support multicultural populations and to prevent paradoxical development of health inequity for those who have difficulties using digital tools to access health information or services [34]. This CBPAR study revealed key barriers and facilitators related to digital health tool uptake among this sample of South Asian Canadians. The findings emphasize the need to overcome language and cultural barriers for meaningful engagement and prioritize community participation to get to the key issues. Future health promotion strategies and research should consider these methods and findings to ensure that community members with different cultural and language backgrounds have the opportunity to inform the development and use of effective, culturally appropriate digital health tools.

Acknowledgments

The authors would like to thank the Vancouver Foundation for funding this research. The authors would also like to thank their partners and partner organizations for their involvement: Canada India Network Society, Progressive Intercultural Community Services Society, Vedic Hindu Cultural Society of British Columbia, Guru Nanak Sikh Gurdwara Society of Delta-Surrey, Fraser Health Authority, British Columbia Ministry of Health Patients as Partners Initiative, and the InterCultural Online Health Network. The authors thank Mr Paul Bains, Dr Gulzar Cheema, Dr Arun Garg, Mr Jay Bains, and Dr Victoria Lee for their advice and involvement. The authors would also like to thank their peer community research team members: Anita Bal, Avneet Brar, Navi Dhaliwal, Jatinder Dhanju, Gunisha Kalra, Sunita (Sonya) Kapoor, Vipin Pandita, and Anisha Takher. The funder had no role in the study design, collection, analysis and interpretation of data, writing of the article, or decision to submit it for publication.

Authors' Contributions

All authors made substantial contributions to the study, either drafted or critically revised this manuscript, gave final approval to publish, and agreed to be held accountable for this work. ES, HM, KA, HNL, and KH contributed to the study conceptualization and design. KH, HNL, and ES contributed to the funding acquisition. ES, HM, KA, and HNL were responsible for the collection of data and project administration. AH, HM, KA, and ES contributed to the data validation and analysis. AH, HM, ES, and KS contributed to the interpretation of results and were responsible for writing the original draft. All authors contributed to revising and finalizing the manuscript.

Conflicts of Interest

None declared.

References

1. Census profile, 2016 census. Statistics Canada. 2017. URL: https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/index.cfm?Lang=E [accessed 2018-01-01]
2. Rana A, de Souza RJ, Kandasamy S, Lear S, Anand SS. Cardiovascular risk among South Asians living in Canada: a systematic review and meta-analysis. CMAJ Open 2014 Jul;2(3):183-191 [FREE Full text] [doi: 10.9778/cmajo.20130064] [Medline: 25295238]
3. Tu JV, Chu A, Rezai MR, Guo H, Maclagan LC, Austin PC, et al. Incidence of major cardiovascular events in immigrants to Ontario, Canada. Circulation 2015 Oct 20;132(16):1549-1559. [doi: 10.1161/circulationaha.115.015345]
4. Sohail QZ, Chu A, Rezai MR, Donovan LR, Ko DT, Tu JV. The risk of ischemic heart disease and stroke among immigrant populations: a systematic review. Can J Cardiol 2015 Sep;31(9):1160-1168. [doi: 10.1016/j.cjca.2015.04.027] [Medline: 26195229]
5. South Asian Health Report. Fraser Health Authority. 2015. URL: https://www.fraserhealth.ca/-/media/Project/FraserHealth/FraserHealth/Health-Topics/South-Asian-Health/201606_South_Asian_Health_Report.pdf [accessed 2018-01-01]
6. Ho K, Novak LH, Stacy E, Jagdis R. InterCultural Online Health Network (iCON): technology enabled wellness support for dementia patients and caregivers. Gerontechnology 2010 [FREE Full text]
7. van Draanen J, Shafique A, Farissi A, Wickramanayake D, Kuttaiya S, Oza S, et al. How to offer culturally relevant type 2 diabetes screening: lessons learned from the South asian diabetes prevention program. Can J Diabetes 2014 Oct;38(5):329-333. [doi: 10.1016/j.jcjd.2013.11.008] [Medline: 24797496]
8. Kalich A, Heinemann L, Ghahari S. A scoping review of immigrant experience of health care access barriers in Canada. J Immigr Minor Health 2016 Jun 21;18(3):697-709. [doi: 10.1007/s10903-015-0237-6] [Medline: 26093784]
9. Braveman P, Gruskin S. Defining equity in health. J Epidemiol Community Health 2003 Apr;57(4):254-258 [FREE Full text] [doi: 10.1136/jech.57.4.254] [Medline: 12646539]
10. Ho K. Key trends in ehealth to propel health transformation and education. UBC Med J 2014;5(2):6-9 [FREE Full text]
11. Oh H, Rizo C, Enkin M, Jadad A. What is eHealth (3): a systematic review of published definitions. J Med Internet Res 2005 Feb 24;7(1):e1 [FREE Full text] [doi: 10.2196/jmir.7.1.e1] [Medline: 15829471]
12. Zibrik L, Khan S, Bangar N, Stacy E, Lauscher H, Ho K. Patient and community centered eHealth: exploring eHealth barriers and facilitators for chronic disease self-management within British Columbia's immigrant Chinese and Punjabi seniors. Health Policy Technol 2015 Dec;4(4):348-356. [doi: 10.1016/j.hlpt.2015.08.002]
13. Valuing time saved: assessing the impact of patient time saved from the adoption of consumer health solutions. The Conference Board of Canada. 2012. URL: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C5&q=%22Valuing+Time+Saved.+Assessing+the+Impact+of+Patient+Time+Saved+from+the+Adoption+of+Consumer+Health+Solutions%22&btnG= [accessed 2021-12-12]
14. Cebul RD, Love TE, Jain AK, Hebert CJ. Electronic health records and quality of diabetes care. N Engl J Med 2011 Sep;365(9):825-833. [doi: 10.1056/nejmsa1102519]
15. Norman C. eHealth literacy 2.0: problems and opportunities with an evolving concept. J Med Internet Res 2011 Dec 23;13(4):e125 [FREE Full text] [doi: 10.2196/jmir.2035] [Medline: 22193243]
16. Ancker J, Miller M, Patel V, Kaushal R, HITEC Investigators. Sociotechnical challenges to developing technologies for patient access to health information exchange data. J Am Med Inform Assoc 2014;21(4):664-670 [FREE Full text] [doi: 10.1136/amiajnl-2013-002073] [Medline: 24064443]
17. Flott K, Callahan R, Darzi A, Mayer E. A patient-centered framework for evaluating digital maturity of health services: a systematic review. J Med Internet Res 2016 Apr 14;18(4):e75 [FREE Full text] [doi: 10.2196/jmir.5047] [Medline: 27080852]
18. Makowsky M, Jones C, Davachi S. Prevalence and predictors of health-related internet and digital device Use in a sample of South Asian adults in Edmonton, Alberta, Canada: results from a 2014 community-based survey. JMIR Public Health Surveill 2021 Jan 08;7(1):e20671 [FREE Full text] [doi: 10.2196/20671] [Medline: 33416506]
19. Klimis H, Thakkar J, Chow CK. Breaking barriers: mobile health interventions for cardiovascular disease. Can J Cardiol 2018 Jul;34(7):905-913. [doi: 10.1016/j.cjca.2018.02.012] [Medline: 29887218]

20. Gordon N, Hornbrook MC. Differences in access to and preferences for using patient portals and other ehealth technologies based on race, ethnicity, and age: a database and survey study of seniors in a large health plan. J Med Internet Res 2016 Mar 04;18(3):e50 [FREE Full text] [doi: 10.2196/jmir.5105] [Medline: 26944212]
21. Olshansky E, Sacco D, Braxter B, Dodge P, Hughes E, Ondeck M, et al. Participatory action research to understand and reduce health disparities. Nurs Outlook 2005 May;53(3):121-126. [doi: 10.1016/j.outlook.2005.03.002] [Medline: 15988448]
22. Wallerstein NB, Duran B. Using community-based participatory research to address health disparities. Health Promot Pract 2006 Jul 30;7(3):312-323. [doi: 10.1177/1524839906289376] [Medline: 16760238]
23. Catalani C, Minkler M. Photovoice: a review of the literature in health and public health. Health Educ Behav 2010 Jun;37(3):424-451. [doi: 10.1177/1090198109342084] [Medline: 19797541]
24. Wang C, Burris MA. Empowerment through photo novella: portraits of participation. Health Educ Q 1994;21(2):171-186. [doi: 10.1177/109019819402100204] [Medline: 8021146]
25. Wang C. Photovoice: a participatory action research strategy applied to women's health. J Womens Health 1999 Mar;8(2):185-192. [doi: 10.1089/jwh.1999.8.185] [Medline: 10100132]
26. Wang C, Burris MA. Photovoice: concept, methodology, and use for participatory needs assessment. Health Educ Behav 1997 Jun;24(3):369-387. [doi: 10.1177/109019819702400309] [Medline: 9158980]
27. 2016 census data. City of Surrey Statistics. 2016. URL: https://www.surrey.ca/files/Neighbourhood-Profile-City-of-Surrey.pdf [accessed 2018-01-01]
28. Leech N, Onwuegbuzie AJ. An array of qualitative data analysis tools: a call for data analysis triangulation. Sch Psychol Q 2007 Dec;22(4):557-584. [doi: 10.1037/1045-3830.22.4.557]
29. Leech N, Onwuegbuzie AJ. Qualitative data analysis: a compendium of techniques and a framework for selection for school psychology research and beyond. Sch Psychol Q 2008 Dec;23(4):587-604. [doi: 10.1037/1045-3830.23.4.587]
30. Onwuegbuzie AJ, Dickinson WB, Leech NL, Zoran AG. A qualitative framework for collecting and analyzing data in focus group research. Int J Qual Methods 2009 Sep 01;8(3):1-21. [doi: 10.1177/160940690900800301]
31. Gudes O, Kendall E, Yigitcanlar T, Pathak V, Baum S. Rethinking health planning: a framework for organising information to underpin collaborative health planning. Health Inf Manag 2010;39(2):18-29. [doi: 10.1177/183335831003900204] [Medline: 20577020]
32. Coombe CM, Chandanabhumma PP, Bhardwaj P, Brush BL, Greene-Moton E, Jensen M, et al. A participatory, mixed methods approach to define and measure partnership synergy in long-standing equity-focused CBPR partnerships. Am J Community Psychol 2020 Dec 03;66(3-4):427-438 [FREE Full text] [doi: 10.1002/ajcp.12447] [Medline: 32744781]
33. Mayan MJ, Daum CH. Worth the risk? Muddled relationships in community-based participatory research. Qual Health Res 2016 Jan 26;26(1):69-76. [doi: 10.1177/1049732315618660] [Medline: 26612889]
34. Weiss D, Rydland H, Øversveen E, Jensen M, Solhaug S, Krokstad SJ. Innovative technologies and social inequalities in health: a scoping review of the literature. PLoS One 2018;13(4):e0195447 [FREE Full text] [doi: 10.1371/journal.pone.0195447] [Medline: 29614114]

Abbreviations

CBPAR: community-based participatory action research
iHEAL: Interactive Health Education Action for Life
PCR: peer community researcher
SDoH: social determinants of health

Edited by G Eysenbach, S Hagens; submitted 23.11.20; peer-reviewed by Y Wang, CO Amaefule, R Halkes; comments to author 30.04.21; revised version received 28.07.21; accepted 22.11.21; published 13.01.22.

Please cite as:
Hyman A, Stacy E, Mohsin H, Atkinson K, Stewart K, Novak Lauscher H, Ho K
Barriers and Facilitators to Accessing Digital Health Tools Faced by South Asian Canadians in Surrey, British Columbia: Community-Based Participatory Action Exploration Using Photovoice
J Med Internet Res 2022;24(1):e25863
URL: https://www.jmir.org/2022/1/e25863
doi:10.2196/25863
PMID:35023842

Original Paper

Using the Social Robot NAO for Emotional Support to Children at a Pediatric Emergency Department: Randomized Clinical Trial

Silvia Rossi[1*], Prof Dr; Silvano Junior Santini[2*], PhD; Daniela Di Genova[2], MD; Gianpaolo Maggi[3], BSc; Alberto Verrotti[4], MD; Giovanni Farello[5], MD; Roberta Romualdi[2], MD; Anna Alisi[6], PhD; Alberto Eugenio Tozzi[7], MD; Clara Balsano[2], MD

[1]Department of Electrical Engineering and Information Technology, University Federico II-Naples, Naples, Italy

[2]Department of Life, Health & Environmental Sciences, School of Emergency and Urgency Medicine, University of L'Aquila, L'Aquila, Italy

[3]Department of Psychology, University of Campania "Luigi Vanvitelli", Caserta, Italy

[4]Department of Pediatrics, University of Perugia, Perugia, Italy

[5]Pediatric Unit, Department of Life, Health & Environmental Sciences, University of L'Aquila, L'Aquila, Italy

[6]Research Unit of Molecular Genetics of Complex Phenotypes, Bambino Gesù Children's Hospital, Istituto di Ricovero e Cura a Carattere Scientifico, Rome, Italy

[7]Multifactorial and Complex Diseases Research Area, Bambino Gesù Children's Hospital, Istituto di Ricovero e Cura a Carattere Scientifico, Rome, Italy

[*]these authors contributed equally

Corresponding Author:
Clara Balsano, MD
Department of Life, Health & Environmental Sciences
School of Emergency and Urgency Medicine
University of L'Aquila
via spennati L'aquila 67100
L'Aquila
Italy
Phone: 39 0862434774
Email: clara.balsano@univaq.it

Abstract

Background: Social robots (SRs) have been used for improving anxiety in children in stressful clinical situations, such as during painful procedures. However, no studies have yet been performed to assess their effect in children while waiting for emergency room consultations.

Objective: This study aims to assess the impact of SRs on managing stress in children waiting for an emergency room procedure through the assessment of salivary cortisol levels.

Methods: This was an open randomized clinical trial in children attending a pediatric emergency department. Children accessing the emergency room were randomized to 1 of 3 groups: (1) playing with a NAO SR, (2) playing with a study nurse, or (3) waiting with parents. The salivary cortisol levels of all children were measured through a swab. Salivary cortisol levels before and after the intervention were compared in the 3 groups. We calculated the effect size of our interventions through the Cohen *d*-based effect size correlation (*r*).

Results: A total of 109 children aged 3-10 years were enrolled in the study, and 94 (86.2%) had complete data for the analyses. Salivary cortisol levels significantly decreased more in the group exposed to robot interaction than in the other two groups (r=0.75). Cortisol levels decreased more in girls (r=0.92) than in boys (r=0.57).

Conclusions: SRs are efficacious in decreasing stress in children accessing the emergency room and may be considered a tool for improving emotional perceptions of children and their families in such a critical setting.

Trial Registration: ClinicalTrials.gov NCT04627909; https://clinicaltrials.gov/ct2/show/study/NCT04627909

(J Med Internet Res 2022;24(1):e29656) doi:10.2196/29656

KEYWORDS

children; emotional health; emergency department; social robots; anxiety; stress

Introduction

Social robots (SRs) may offer a multifactorial sensory experience to children and distract them from stressful situations, as it frequently happens during health care encounters [1,2]. SRs are designed to interact and communicate with human beings by play, gestures, poses, gaze, and colors and have been successfully used with pediatric patients in different settings [3,4], although not all the available devices on the market have anthropomorphic, physical, and behavioral qualities to establish a virtuous collaboration with children [5-8].

Addressing the emotional needs of children who present to the pediatric emergency department (ED) is a complex task that requires the management of anxiety and pain during medical procedures [9]. In such situations, emotional stress may affect the outcome of emergency interventions due to the lack of cooperation of young patients with health care providers, may delay the diagnosis, and may prolong medical procedures [10,11]. For this reason, interventions for reducing stress in children attending the ED are highly desirable [12-14].

Although several studies have explored the impact of SRs on negative emotions in children [3,15,16], their efficacy in reducing stress in hospitalized children is still uncertain. Moreover, to the best of our knowledge, the impact of SRs on stress in children accessing the ED has not been investigated yet.

We therefore conducted a randomized clinical trial with the aim of comparing SR interaction with playing with a nurse, or no intervention, in children accessing the emergency room prior to entering the medical office. Our hypothesis was that SR interaction is superior to other interventions in reducing stress. As the biological response in stressful situations includes the activation of the hypothalamic-pituitary-adrenal axis, a glucocorticoid response and subsequent cortisol release [17], we measured as an outcome salivary cortisol levels in children.

Methods

Study Design and Participants

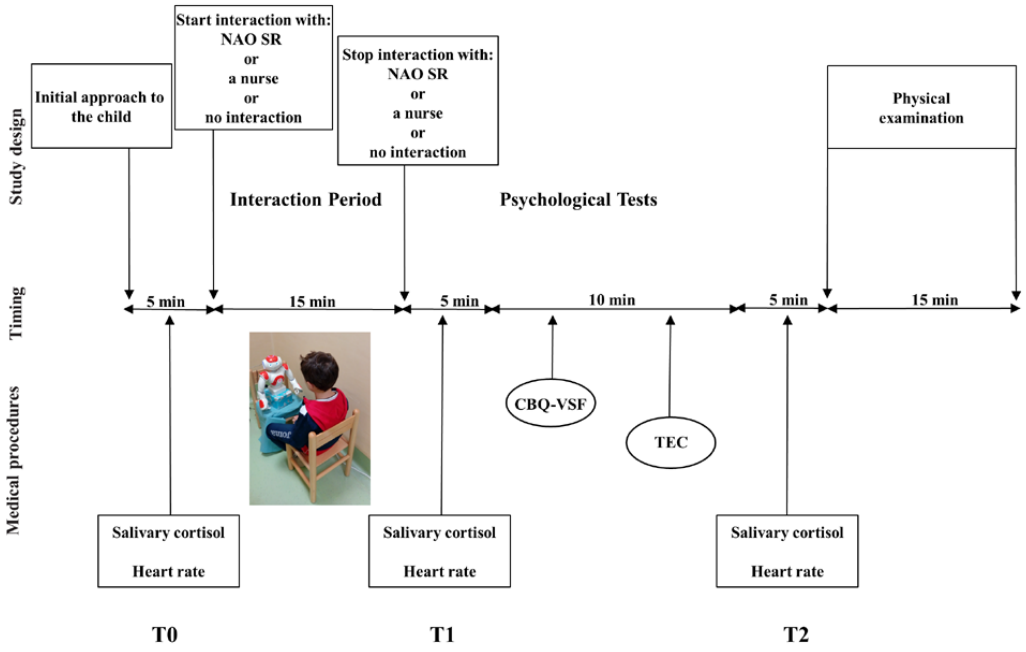
This was an open, 3-arm, parallel, randomized clinical trial conducted on children who attended the pediatric ED of the San Salvatore Hospital of L'Aquila from September 1, 2019, to February 29, 2020. Children eligible for the trial were those 3-10 years old, who were assigned a white (not critical), green (not very critical), or yellow (moderately critical) code with no neurological condition at triage, as described by Piccotti et al [18], and who could safely wait for nonurgent care in the waiting room. Exclusion criteria were age range <3 and >10 years; parents not fluent in Italian; yellow code for headache due to recent trauma with visual disturbances, headache accompanied by neck rigidity, vomiting, or indifference to the environment; presence of dyspnea; significant trauma to the head with an altered state of consciousness; and red code (very critical).

Parents and children who attended the pediatric ED of the San Salvatore Hospital of L'Aquila in the study period were invited to participate in the study, and informed consent (parent) and assent (child) were obtained. Enrolment was restricted between 8:30 AM and 10:30 AM to avoid the differences in cortisol levels due to their circadian fluctuation. The children were then randomized to 1 of 3 groups: (1) interaction with an SR, (2) playing with a study nurse, or (3) a control group assigned to routine waiting with parents. For all the 3 groups, there were no restrictions regarding whether the parents could stay with their children; however, only children in the control group could interact with them. All children in the study received standard care.

Once randomized, the children's temperature, heart rate, and a salivary sample were taken (T0). The children were then exposed to 1 of the 3 interventions for 15 minutes. Immediately after the intervention, the children's heart rate was measured again and an additional salivary sample was taken (T1). After that, the children underwent a battery of psychological tests (see the Interventions section), which took nearly 10 minutes to complete, and a third heart rate measure and a salivary sample were taken thereafter (T2). Finally, the children entered the medical office in the ED. The entire process is illustrated in Figure 1.

The primary outcome measure of the study was the difference in cortisol level of patients before (T0) and after the intervention (T2).

Figure 1. Cohort assembly: study procedure and timing of outcome assessment. CBQ-VSF: Children's Behavior Questionnaire-Very Short Form; T0-T1-T2: timing of cortisol level measurement; TEC: Test of Emotion Comprehension.



Interventions

The NAO robot is a small SR, developed by Softbank Robotics, programmed to interact with children (Figure 1) [10,11]. The robot uses cognitive-behavioral distraction strategies appropriate to age (songs, stories, jokes, games, riddles) with children [19-22]. The robot started asking the child's name and age. The robot then autonomously selected the appropriate interaction according to age. Questions posed by the robot to tailor interactions included school attendance, ability to count, favorite subjects, and cartoons. Based on this information, the robot asked the child to count together and to talk about school and favorite subjects and cartoons. A further interpretation of the answers provided by children allowed the robot to make comments and to play the theme songs of the cartoon. The robot also played songs and jokes appropriate for the children's age, inviting them to guess the title or the solution. An additional interaction was about nursery rhymes and tongue twisters. Finally, the robot played guessing games about animal noises and other riddles.

The NAO robot was equipped with natural language processing technologies to understand the child's speech. However, in the case of background noise or incorrect pronunciation, a doctor was present to enter manually answers into the NAO software by the use of a hidden personal computer.

In the second intervention group, children played with nurses by coloring children's books or using toys available in the waiting room.

Children in the control group stayed with their parents and were free to play while waiting.

Salivary Analysis and Heart Rate Measurement

Nurses who were not directly involved in this study measured the children's heart rate and temperature.

Salivary samples were collected after rinsing the mouth with water to prevent food contamination. A swab from a Salivette device (cat. 51.1534, Sarstedt, Nümbrecht, Germany) was placed under the tongue of the participant, and saliva was absorbed for 60 seconds. The saliva-saturated Salivette swab then was placed in a polypropylene tube and centrifuged at 1000 ×*g* for 15 minutes at 4°C for saliva extraction. The saliva sample was then frozen at −20°C. To start the analysis of the salivary cortisol levels, the saliva was centrifuged at 2500 ×*g* for 20 minutes after thawing, and the clear supernatant was used in the analysis.

The salivary cortisol levels were assessed using a commercially available DetectX Cortisol Enzyme Immunoassay Kit (cat. K003-H1W, Arbor Assays, MI, USA) and a Victor3 microplate reader (PerkinElmer, Waltham, MA, USA) according to the manufacturers' instructions.

Psychological Tests

To verify that salivary cortisol levels may have been not confounded by temperament or emotional management, we performed a series of psychological tests at T1. Although this evaluation was performed after the intervention to avoid interferences with the study and the ED workflow, it is important to ensure that the tests performed at this time are not influenced by external stimuli, including those of the intervention. For this reason, these tests were considered useful for comparing participants at baseline. The evaluation consisted of the Children's Behavior Questionnaire-Very Short Form (CBQ-VSF), which evaluates children's temperament, and the Test of Emotion Comprehension (TEC), which were

administered and completed in the presence of assistant psychologists.

The CBQ-VSF [23,24] is a 36-item questionnaire that is completed by the child caregiver to assess the temperament of children aged 3-8 years. It is designed to measure 3 broad dimensions: surgency/extraversion, negative affectivity, and effortful control. The surgency/extraversion scale is characterized by high activity levels, high-intensity pleasure seeking, and low shyness and impulsivity. The second dimension, negative affectivity, is defined by feelings of sadness, discomfort, frustration, and fear. The effortful control scale encompasses inhibitory control, attentional focus, low-intensity pleasure, and perceptual sensitivity [25]. Caregivers were asked to rate how well the items describe the child's reaction in a variety of situations. The responses were given on a 7-point scale ranging from 1 (extremely untrue of my child) to 7 (extremely true of my child).

The TEC [26] evaluates the understanding and managing of emotions in children aged 3-11 years. It consists of 9 components, namely the ability of recognition of emotions based on facial expressions (labeling), the comprehension of external emotional causes, the impact of desire on emotions, emotions based on beliefs, the influence of memories on emotions, the possibility of emotional regulation, the possibility of hiding an emotional state, having mixed emotions, and the contribution of morality to emotional experiences [27].

Statistical Analysis and Sample Size

Randomization of the intervention was made through computer-generated randomization codes (Random Allocation Software version 1.0, Isfahan University of Medical Sciences) by an independent researcher, and an envelope containing sequential numbers was given to the parents.

Data for all variables were checked for normal distribution through the Shapiro-Wilk test. As the test indicated that distributions were not normally distributed, we used medians and IQRs for continuous variables. To evaluate the differences in cortisol levels and heart rate at the 3 different time points within each group, we used the Friedman nonparametric test. Post hoc analyses with the Wilcoxon signed-rank test were carried out to determine where the observed differences were. To evaluate the strength of the relationship between the variables of our interest, we calculated the effect size of our interventions through the Cohen *d*-based effect size correlation (r).

Linear regression analyses were carried out to evaluate the relationship between salivary cortisol levels and heart rate registered at each time within each age group.

Sample size was calculated based on variations of cortisol levels. It was estimated that to detect a reduction in cortisol levels of at least 0.6 ng/mL from T0 to T2, assuming an SD of 0.17 for cortisol levels, with a 95% statistical power (CI, Z=1.96; proportion of the population, π=0.80; margin of error E=5%, prevalence control group, P_0=.85, prevalence health care personnel group, P_1=.75, NAO robot group, P_2=.80), we evaluated 19 children in each group.

All statistical analyses were performed using Graphpad Prism version 9.0.

The trial was approved by the local ethics committee of the San Salvatore Hospital of L'Aquila (IRB protocol no. 2666 06-25-2019) and registered on ClinicalTrials.gov (identification no. NCT04627909).

Results

Demographic, Psychological, and Other Baseline Characteristics

We invited 145 consecutive families whose children were eligible to participate in the study, and we enrolled a total of 109 patients. Reasons for refusing participation were (1) parents did not want to let their children interact with NAO (n=15, 42%), parents thought that their children were already overexposed to electronic devices (n=8, 22%), parents were particularly concerned about their child's illness (n=7, 19%); and consent by both parents could not be obtained (n=6, 17%).

Of the 109 children, 94 (86.2%) were included in the final analysis as 5 salivary samples in each of the 3 groups could not be analyzed because of sampling problems. Among them, 32 (34%) were in the NAO robot group, 31 (33%) were in the group playing with a study nurse, and 31 (33%) remained with their parents (Figure 2).

A description of demographic, psychological, and other baseline characteristics of children by intervention group are reported in Tables 1 and 2 and in Multimedia Appendices 1 and 2. The baseline characteristics of children in the 3 groups were balanced and did not show significant differences.

Of all 94 participants, 48 (51%) were boys. There were no significant differences in the salivary cortisol levels, heart rate, and temperature among the groups at T0 (Table 1). The average total score of the TEC for children interacting with the NAO robot was 6.8 (2.0), while it was 6.5 (1.2) in the group of children playing with nurses and 6.3 (2.0) in the control group (Table 2). The temperaments of the 3 groups evaluated using the CBQ-VSF displayed similar profiles for extraversion/surgency, negative affectivity, and effortful control; the distribution of the scores is reported in Table 2. These results were obtained in 71 of 94 (76%) children only, because 23 children (24%) did not complete the tests due to ED time constraints.

The lowest cortisol levels at T2 were measured in children with NAO robot interaction (Figure 3A). In this group, cortisol levels at T2 significantly differed from those at T0 (3.50 vs 1.87 ng/mL, P<.001; Figure 3A). In addition, cortisol levels slightly decreased at T2 in children playing with nurses, and they were similar at T0 and T1 in all groups.

When exploring differences in salivary cortisol levels by sex, we found a more pronounced decrease at T2 in girls interacting with the SR, than in boys (Figure 3A-C). The trend was different in boys playing with a nurse, who showed an increase in cortisol levels at T1, which returned to levels similar to baseline at T2, while no significant variations were observed in the control group (Figure 3B).

When looking at the trends in the heart rate overall, we found the same differences observed for salivary cortisol levels (Figure 4A). Moreover, regression analysis showed a significant linear correlation between salivary cortisol levels and heart rate in all intervention groups (Figure 4B-D).

Figure 2. Enrollment and randomization of children.

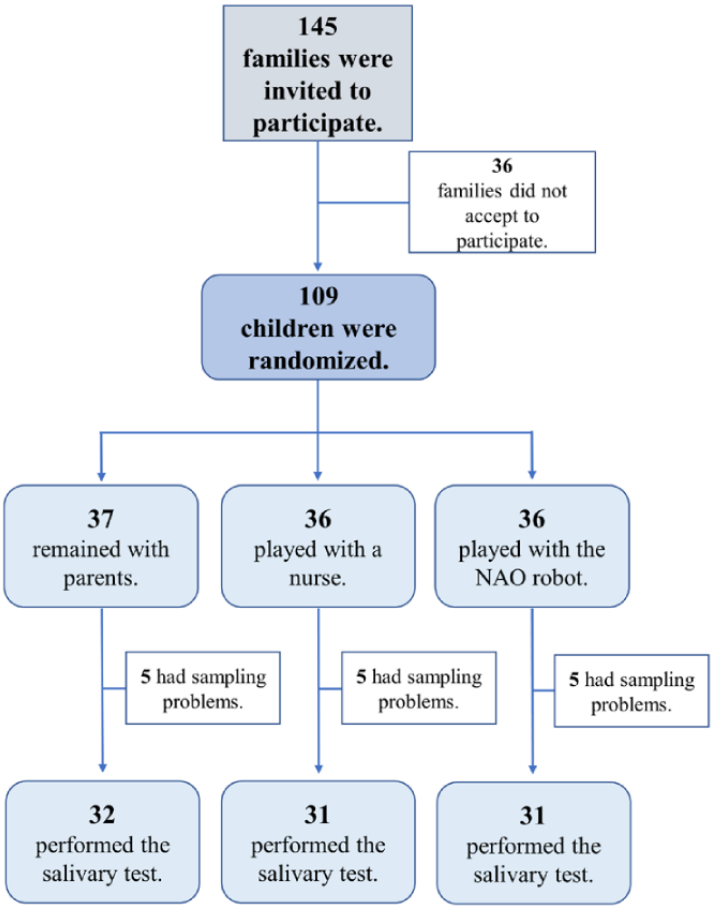


Table 1. Demographics by group.

Demographics	All groups (N=94)	Control group (n=31)	Study nurse group (n=31)	NAO robot group (n=32)
Age, median (IQR)	7 (5-8)	7 (6-7.5)	7 (5.5-8)	6 (5-7)
Sex, n (%)				
Boys	48 (51.1)	16 (51.6)	15 (48.4)	17 (53.1)
Girls	46 (48.9)	15 (48.4)	16 (51.6)	15 (46.9)
Triage code, n (%)				
White	0	0	0	0
Green	90 (95.7)	30 (96.8)	30 (96.8)	30 (93.7)
Yellow	4 (4.3)	1 (3.2)	1 (3.2)	2 (6.3)
Discharge, n (%)				
Home discharge	91 (96.8)	30 (96.8)	30 (96.8)	31 (96.9)
Hospitalization	3 (3.2)	1 (3.2)	1 (3.2)	1 (3.1)
T0 heart rate (bpm), median (IQR)	103 (97.25-117)	102 (95-116)	104 (98-114.5)	105 (99-123.5)
T0 cortisol levels (ng/mL), median (IQR)	3.54 (3.26-3.80)	3.65 (3.35-3.89)	3.41 (3.13-3.67)	3.51 (3.33-3.80)
Temperature (°C), median (IQR)	36.4 (36-36.9)	36.5 (36.2-36.9)	36.2 (36.0-36.5)	36.5 (36-36.9)

Table 2. Psychological variables by group.

Psychological variables	Total sample (N=71)	Control group (n=19)	Study nurse group (n=23)	NAO robot group (n=29)
TEC[a], mean (SD)				
Total score	6.6 (1.8)	6.3 (2.0)	6.5 (1.2)	6.8 (2.0)
Component[b] I	1.0 (0.2)	1.0 (0.0)	1.0 (0.2)	1.0 (0.2)
Component II	0.9 (0.3)	0.8 (0.4)	0.9 (0.3)	0.9 (0.3)
Component III	0.9 (0.3)	0.9 (0.3)	0.8 (0.4)	0.9 (0.3)
Component IV	0.7 (0.5)	0.6 (0.5)	0.7 (0.5)	0.8 (0.4)
Component V	0.8 (0.4)	0.6 (0.5)	0.9 (0.3)	0.8 (0.4)
Component VI	0.5 (0.5)	0.4 (0.5)	0.4 (0.5)	0.5 (0.5)
Component VII	0.7 (0.4)	0.8 (0.4)	0.7 (0.5)	0.7 (0.5)
Component VIII	0.4 (0.5)	0.4 (0.5)	0.3 (0.5)	0.5 (0.5)
Component IX	0.7 (0.4)	0.7 (0.5)	0.8)0.4)	0.8)0.4)
CBQ-VSF[c], mean (SD)				
Extraversion/surgency	52.8 (8.4)	57.3 (6.3)	49.7 (9.8)	53.1 (7.5)
Negative affectivity	51.0 (8.0)	55.6 (7.0)	51.6 (6.8)	49.1 (8.6)
Effortful control	66.2 (8.1)	61.2 (8.1)	65.4 (8.1)	68.3 (7.4)

[a]TEC: Test of Emotion Comprehension.

[b]Component I: recognition of emotions; component II: comprehension of external emotional causes; component III: impact of desire on emotions; component IV: emotions based on beliefs; component V: memory influence on emotions; component VI: possibility of emotional regulation; component VII: possibility of hiding an emotional state; component VIII: mixed emotions; component IX: contribution of morality to emotional experiences.

[c]CBQ-VSF: Children's Behavior Questionnaire-Very Short Form.

Figure 3. Salivary cortisol levels (A) in the whole sample, (B) in boys, and (C) in girls. Data are expressed as the median and IQR (*P<.05, **P<.01, ***P<.001). r represents the Cohen d-based effect size correlation.

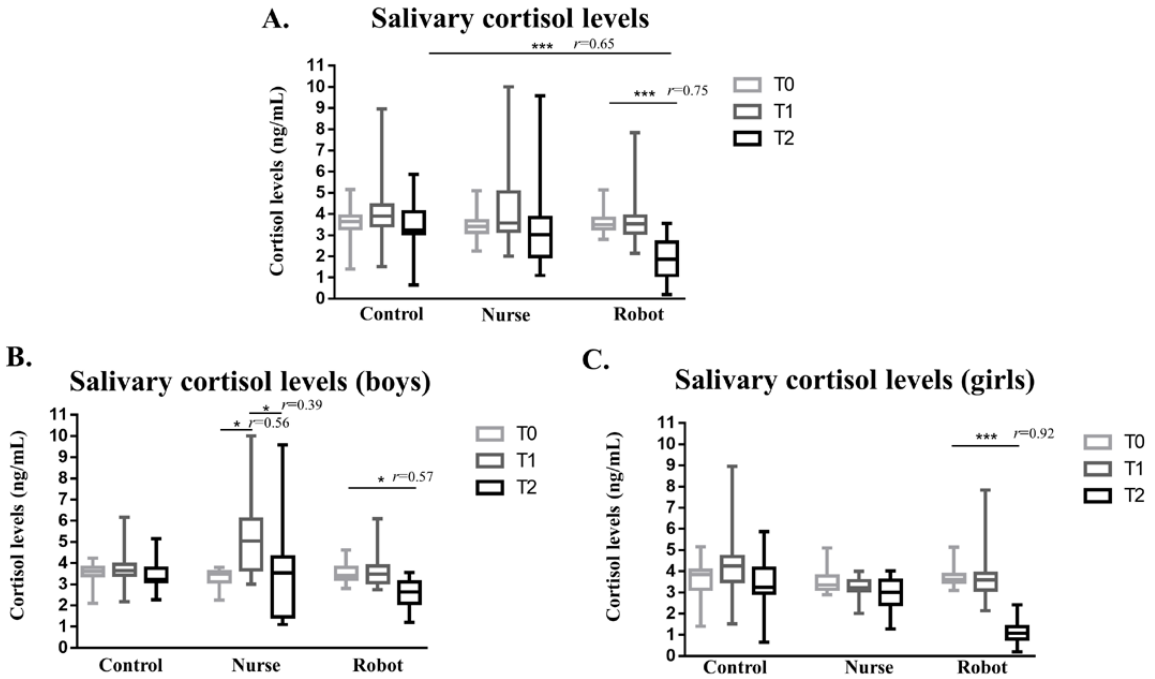
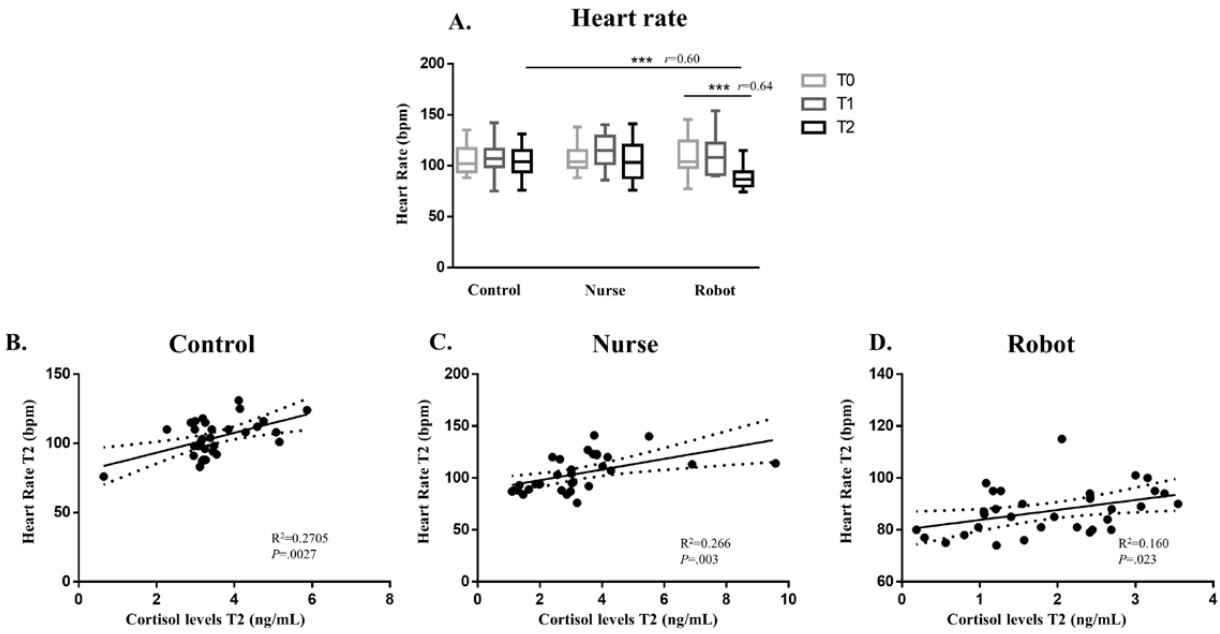


Figure 4. Heart rate (A) in the whole sample. Data are expressed as the median and IQR (***P<.001; r represents the Cohen d-based effect size correlation). Direct linear regression between heart rate and salivary cortisol levels at T2 in the control group (B), study nurse group (C), and NAO robot group (D).



Discussion

Principal Findings

Our study shows that playing with an SR while waiting for a medical procedure in the emergency room may decrease the stress level in children, as demonstrated by a decrease in salivary cortisol levels. Our study shows that this effect is visible not immediately after the intervention but nearly 20 minutes after robot interaction, according to the physiological cortisol response to stressor stimuli changes that occur 15-30 minutes after stressor onset [28].

Moreover, the decrease in cortisol levels is more evident in girls than in boys.

Sex differences in the attitude to interacting with SRs have been already reported, and a previous study conducted in Japan suggested that females have more negative attitudes than males toward robot interaction [29,30]. Although our findings are in contrast with these observations, this is not surprising as cultural background, setting, and perceived stereotypes may affect the attitude of children in using technology [29,31-34].

A recent systematic review examined the effect of SRs on anxiety and distress in children, and 4 of the 10 (40%) included studies used the NAO robot, the same device as in our study [3]. SRs reduced distress and anxiety in all reviewed studies, although the quality of evidence was low mainly due to the small sample size. Our results are in line with the published literature; however, we are not aware of any study on the effect of SRs in children conducted in the emergency room setting. Moreover, differently from the available studies in the literature, we used salivary cortisol levels as an outcome, as they may be measured through an easy and noninvasive method [35-37] and are associated with stress from nonpharmacological interventions [36,38-40].

We administered psychological tests to assess the temperament of children that might affect the control of anxiety in distressing situations. Higher scores on surgency/extraversion temperament have an impact on the cortisol production independently from the assigned intervention arm, demonstrating that children with more extroverted personalities have less emotional distress at baseline. These findings further support the idea that emotion regulation abilities contribute to reduced fear and anxiety in distressing situations [37,38,41,42].

More than 15% (6/36) of parents of children, eligible for our study, refused to participate in the study because they were concerned about exposure of their children to electronic devices. Although setting and cultural background may play a role, this observation should be considered if strategies including SRs for reducing stress and anxiety will be put in place in routine clinical practice.

Strengths and Limitations

This study had several strengths. Its design prevented common confounding effects, as shown by the baseline characteristics of participants that were similar across intervention groups. Moreover, we used salivary cortisol levels to assess the stress level in children, which represents an objective measure of this outcome. To validate our results, we also studied the correlation between salivary cortisol levels and heart rate.

A limitation was that our study was conducted in a single clinical center and we did not administer any psychological test to assess stress/anxiety to avoid slowing down medical procedures in the emergency room. However, psychological questionnaires are

less reliable than salivary cortisol levels in detecting slight differences in children's stress levels.

Although SRs represent a promising tool in managing children's distress in the emergency room, the cultural background and setting may strongly affect the acceptance of these devices. However, we frequently noticed positive comments from parents, such as "My son didn't want to leave the NAO robot," "My son wanted to take the robot with him," "My son told me that the NAO robot was amazing," and "My son asked, for the first time, to come back again to the ED." Additional studies may better address determinants of acceptability in different settings.

Conclusion

In conclusion, our study demonstrates that SRs are effective in decreasing the stress of children in health care emergency contexts. These devices may be integrated in the pediatric ED workflow, with benefits for patients and families and potentially to speed up clinical procedures. To this aim, future and larger studies in different settings should be promoted. Moreover, in circumstances where social contacts should be prevented, such as during the COVID-19 pandemic, SRs may play an important role in improving the emotional experiences of children and their families, and disease outcomes.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Psychological features of children. (A) TEC and (B-D) tests. Data are shown as the mean and SD. Statistically significant differences in the TEC and CBQ scores between groups were not detected. CBQ: Children's Behavior Questionnaire; TEC: Test of Emotion Comprehension.
[PNG File , 46 KB - jmir_v24i1e29656_app1.png]

Multimedia Appendix 2
Vital signs of children. Heart rate (A) and temperature (B) at T0. Data are expressed as the median and IQR. No statistically significant differences in the TEC and CBQ between groups were detected. CBQ: Children's Behavior Questionnaire; TEC: Test of Emotion Comprehension.
[PNG File , 26 KB - jmir_v24i1e29656_app2.png]

Multimedia Appendix 3
CONSORT-eHEALTH checklist (V 1.6.1).
[PDF File (Adobe PDF File), 1249 KB - jmir_v24i1e29656_app3.pdf]

References

1. DeMore M, Cohen L. Distraction for pediatric immunization pain: a critical review. J Clin Psychol Med Settings 2005 Dec;12(4):281-291. [doi: 10.1007/s10880-005-7813-1]
2. Beran TN, Ramirez-Serrano A, Vanderkooi OG, Kuhn S. Humanoid robotics in health care: an exploration of children's and parents' emotional reactions. J Health Psychol 2015 Jul;20(7):984-989. [doi: 10.1177/1359105313504794] [Medline: 24140615]
3. Littler B, Alessa T, Dimitri P, Smith C, de Witte L. Reducing negative emotions in children using social robots: systematic review. Arch Dis Child 2021 Nov;106(11):1095-1101. [doi: 10.1136/archdischild-2020-320721] [Medline: 33685936]
4. Rossi S, Larafa M, Ruocco M. Emotional and behavioural distraction by a social robot for children anxiety reduction during vaccination. Int J of Soc Robotics 2020 Jan 02;12(3):765-777. [doi: 10.1007/s12369-019-00616-w]
5. Meghdari A, Shariati A, Alemi M, Vossoughi GR, Eydi A, Ahmadi E, et al. Arash: a social robot buddy to support children with cancer in a hospital environment. Proc Inst Mech Eng H 2018 Jun;232(6):605-618. [doi: 10.1177/0954411918777520] [Medline: 29890934]
6. Moerman CJ, Jansens RM. Using social robot PLEO to enhance the well-being of hospitalised children. J Child Health Care 2021 Sep 25;25(3):412-426. [doi: 10.1177/1367493520947503] [Medline: 32840383]
7. Ferrante G, Vitale G, Licari A, Montalbano L, Pilato G, Infantino I, et al. Social robots and therapeutic adherence: a new challenge in pediatric asthma? Paediatr Respir Rev 2021 Dec 05;40:46-51. [doi: 10.1016/j.prrv.2020.11.001] [Medline: 33386280]
8. Scassellati B, Admoni H, Matarić M. Robots for use in autism research. Annu Rev Biomed Eng 2012;14:275-294. [doi: 10.1146/annurev-bioeng-071811-150036] [Medline: 22577778]
9. Logan DE, Breazeal C, Goodwin MS, Jeong S, O'Connell B, Smith-Freedman D, et al. Social robots for hospitalized children. Pediatrics 2019 Jul;144(1):e20181511 [FREE Full text] [doi: 10.1542/peds.2018-1511] [Medline: 31243158]
10. Lerwick JL. Minimizing pediatric healthcare-induced anxiety and trauma. World J Clin Pediatr 2016 May 08;5(2):143-150 [FREE Full text] [doi: 10.5409/wjcp.v5.i2.143] [Medline: 27170924]

11. Roohafza H, Kabir A, Sadeghi M, Shokouh P, Ahmadzad-Asl M, Khadem-Maboudi AA, et al. Stress as a risk factor for noncompliance with treatment regimens in patients with diabetes and hypertension. ARYA Atheroscler 2016 Jul;12(4):166-171 [FREE Full text] [Medline: 28149311]
12. MacLean S, Obispo J, Young KD. The gap between pediatric emergency department procedural pain management treatments available and actual practice. Pediatr Emerg Care 2007 Feb;23(2):87-93. [doi: 10.1097/PEC.0b013e31803] [Medline: 17351407]
13. Young KD. Pediatric procedural pain. Ann Emerg Med 2005 Feb;45(2):160-171. [doi: 10.1016/j.annemergmed.2004.09.019] [Medline: 15671974]
14. Zempsky WT, Cravero JP, American Academy of Pediatrics Committee on Pediatric Emergency MedicineSection on AnesthesiologyPain Medicine. Relief of pain and anxiety in pediatric patients in emergency medical systems. Pediatrics 2004 Nov;114(5):1348-1356. [doi: 10.1542/peds.2004-1752] [Medline: 15520120]
15. Okita SY. Self-other's perspective taking: the use of therapeutic robot companions as social agents for reducing pain and anxiety in pediatric patients. Cyberpsychol Behav Soc Netw 2013 Jun;16(6):436-441. [doi: 10.1089/cyber.2012.0513] [Medline: 23505968]
16. Trost M, Ford A, Kysh L, Gold J, Matarić M. Socially assistive robots for helping pediatric distress and pain: a review of current evidence and recommendations for future research and practice. Clin J Pain 2019 May;35(5):451-458 [FREE Full text] [doi: 10.1097/AJP.0000000000000688] [Medline: 30951515]
17. Smith SM, Vale WW. The role of the hypothalamic-pituitary-adrenal axis in neuroendocrine responses to stress. Dialogues Clin Neurosci 2006;8(4):383-395 [FREE Full text] [Medline: 17290797]
18. Piccotti E, Magnani M, Tubino B, Sartini M, Di Pietro P. Assessment of the triage system in a pediatric emergency department. A pilot study on critical codes. J Prev Med Hyg 2008 Sep;49(3):120-123. [Medline: 19278139]
19. Cohen LL. Reducing infant immunization distress through distraction. Health Psychol 2002 Mar;21(2):207-211. [Medline: 11950112]
20. Mahony DL, Burroughs WJ, Hieatt AC. The effects of laughter on discomfort thresholds: does expectation become reality? J Gen Psychol 2001 Apr;128(2):217-226. [doi: 10.1080/00221300109598909] [Medline: 11506050]
21. Whitehead-Pleaux AM, Baryza MJ, Sheridan RL. The effects of music therapy on pediatric patients' pain and anxiety during donor site dressing change. J Music Ther 2006;43(2):136-153. [doi: 10.1093/jmt/43.2.136] [Medline: 16897906]
22. Stuber M, Hilber SD, Mintzer LL, Castaneda M, Glover D, Zeltzer L. Laughter, humor and pain perception in children: a pilot study. Evid Based Complement Alternat Med 2009 Jun;6(2):271-276 [FREE Full text] [doi: 10.1093/ecam/nem097] [Medline: 18955244]
23. Rothbart MK, Ahadi SA, Hershey KL, Fisher P. Investigations of temperament at three to seven years: the Children's Behavior Questionnaire. Child Dev 2001;72(5):1394-1408. [doi: 10.1111/1467-8624.00355] [Medline: 11699677]
24. Putnam SP, Rothbart MK. Development of short and very short forms of the Children's Behavior Questionnaire. J Pers Assess 2006 Aug;87(1):102-112. [doi: 10.1207/s15327752jpa8701_09] [Medline: 16856791]
25. Rothbart MK, Putnam SP. Temperament and socialization. In: Pulkkinen L, Caspi A, editors. Paths to Successful Development: Personality in the Life Course. UK: Cambridge University Press; 2002:19-45.
26. Pons F, Harris P, de Rosnay M. Emotion comprehension between 3 and 11 years: developmental periods and hierarchical organization. Eur J Dev Psychol 2004 Jun;1(2):127-152. [doi: 10.1080/17405620344000022]
27. Rocha AA, Roazzi A, Lopes da Silva A, Candeias AA, Minervino CA, Roazzi MM, et al. Test of Emotion Comprehension: exploring the underlying structure through confirmatory factor analysis and similarity structure analysis. In: Roazzi A, de Souza BC, Bilsky W, editors. Facet Theory: Searching for Structure in Complex Social, Cultural and Psychological Phenomena. Recife, Brazil: Editora UFPE; Dec 15, 1975:66-84.
28. Dickerson S, Kemeny M. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull 2004 May;130(3):355-391. [doi: 10.1037/0033-2909.130.3.355] [Medline: 15122924]
29. Nomura T. Robots and Gender. Gend Genome 2017 Mar 01;1(1):18-26. [doi: 10.1089/gg.2016.29002.nom]
30. Nomura T, Suzuki T, Kanda T, Kato K. Measurement of negative attitudes toward robots. IS 2006 Nov 15;7(3):437-454. [doi: 10.1075/is.7.3.14nom]
31. Tung F. Influence of gender and age on the attitudes of children towards humanoid robots. 2011 Presented at: HCI International 2011; July 9-14, 2011; Orlando, FL, USA p. 637-646. [doi: 10.1007/978-3-642-21619-0_76]
32. Sandygulova A, O'Hare G. Investigating the impact of gender segregation within observational pretend play interaction. 2016 Presented at: ACM/IEEE International Conference on Human-Robot Interaction; 2016; Christchurch, New Zealand p. 399-406. [doi: 10.1109/hri.2016.7451778]
33. Essau CA, Sakano Y, Ishikawa S, Sasagawa S. Anxiety symptoms in Japanese and in German children. Behav Res Ther 2004 May;42(5):601-612. [doi: 10.1016/S0005-7967(03)00164-5] [Medline: 15033504]
34. Hosseini L, Khazali H. Procedia - Soc Behav Sci 2013 Jul;84:41-46. [doi: 10.1016/j.sbspro.2013.06.506]
35. Nicolson N. Measurement of cortisol. In: Handbook of Physiological Research Methods in Health Psychology. NY: Sage Publications; 2008:37-38.

36. Estrada C, Cuenca L, Cano-Fernandez L, Gil-Martinez AL, Sanchez-Rodrigo C, González-Cuello AM, et al. Voluntary exercise reduces plasma cortisol levels and improves transitory memory impairment in young and aged Octodon degus. Behav Brain Res 2019 Nov 05;373:112066. [doi: 10.1016/j.bbr.2019.112066] [Medline: 31269420]
37. Miller R, Stalder T, Jarczok M, Almeida DM, Badrick E, Bartels M, et al. The CIRCORT database: reference ranges and seasonal changes in diurnal salivary cortisol derived from a meta-dataset comprised of 15 field studies. Psychoneuroendocrinology 2016 Nov;73:16-23 [FREE Full text] [doi: 10.1016/j.psyneuen.2016.07.201] [Medline: 27448524]
38. Thirthalli J, Naveen G, Rao M, Varambally S, Christopher R, Gangadhar B. Cortisol and antidepressant effects of yoga. Indian J Psychiatry 2013 Jul;55(Suppl 3):S405-S408 [FREE Full text] [doi: 10.4103/0019-5545.116315] [Medline: 24049209]
39. de la Rubia Ortí JE, García-Pardo MP, Iranzo CC, Madrigal JJC, Castillo SS, Rochina MJ, et al. Does music therapy improve anxiety and depression in Alzheimer's patients? J Altern Complement Med 2018 Jan;24(1):33-36. [doi: 10.1089/acm.2016.0346] [Medline: 28714736]
40. Antonelli M, Barbieri G, Donelli D. Effects of forest bathing (shinrin-yoku) on levels of cortisol as a stress biomarker: a systematic review and meta-analysis. Int J Biometeorol 2019 Aug;63(8):1117-1134. [doi: 10.1007/s00484-019-01717-x] [Medline: 31001682]
41. Rossi S, Conti D, Garramone F, Santangelo G, Staffa M, Varrasi S, et al. The role of personality factors and empathy in the acceptance and performance of a social robot for psychometric evaluations. Robotics 2020 May 23;9(2):39. [doi: 10.3390/robotics9020039]
42. Cisler JM, Olatunji BO, Feldner MT, Forsyth JP. Emotion regulation and the anxiety disorders: an integrative review. J Psychopathol Behav Assess 2010 Mar;32(1):68-82 [FREE Full text] [doi: 10.1007/s10862-009-9161-1] [Medline: 20622981]

Abbreviations

CBQ-VSF: Children's Behavior Questionnaire-Very Short Form
ED: emergency department
SR: social robot
TEC: Test of Emotion Comprehension

Edited by R Kukafka; submitted 15.04.21; peer-reviewed by O Korn, C Sutherland; comments to author 01.06.21; revised version received 11.06.21; accepted 09.11.21; published 13.01.22.

Please cite as:
Rossi S, Santini SJ, Di Genova D, Maggi G, Verrotti A, Farello G, Romualdi R, Alisi A, Tozzi AE, Balsano C
Using the Social Robot NAO for Emotional Support to Children at a Pediatric Emergency Department: Randomized Clinical Trial
J Med Internet Res 2022;24(1):e29656
URL: https://www.jmir.org/2022/1/e29656
doi:10.2196/29656
PMID:34854814

Original Paper

Implementation of the Flexible Assertive Community Treatment (FACT) Model in Norway: eHealth Assessment Study

Erlend Bønes[1,2], MSc; Conceição Granja[1,2], PhD; Terje Solvoll[1,2], PhD

[1]Norwegian Centre for e-Health Research, University Hospital of North Norway, Tromsø, Norway

[2]Faculty of Nursing and Health Sciences, Nord University, Bodø, Norway

Corresponding Author:
Erlend Bønes, MSc
Norwegian Centre for e-Health Research
University Hospital of North Norway
PO Box 35
Tromsø, 9038
Norway
Phone: 47 97655680
Email: erlend.bones@ehealthresearch.no

Abstract

Background: Flexible Assertive Community Treatment (FACT) is a model for treatment of long-term severe mental disorders. This method has become more widespread in Norway.

Objective: The objective of our study was to examine how the implementation of FACT teams in Norway has been affected by eHealth policy, infrastructure, and regulations. Another objective was to examine existing literature on eHealth interventions and challenges within FACT teams.

Methods: We have examined Norwegian policy regulating mental health services, laws and regulations, eHealth infrastructure, relevant literature on FACT teams, and current implementation of FACT in Norway.

Results: FACT teams are a wanted part of the Norwegian service system, but the current eHealth infrastructure makes sharing of data within teams and levels of health care challenging, even if eHealth regulations allow such sharing. This has been shown to be an issue in the current implementation of FACT teams in Norway. There is little or no existing research on the eHealth challenges facing FACT teams.

Conclusions: Weaknesses in the Norwegian eHealth infrastructure have been a barrier for an easy implementation of FACT teams in Norway. It is difficult to share information between the different levels of health care. We need systems that allow for easy, secure sharing of health information to and between the FACT team members and other involved health care workers.

(J Med Internet Res 2022;24(1):e32220) doi:10.2196/32220

KEYWORDS

mental health; FACT; electronic health records; eHealth; FACT implementation; EHR; implementation; assessment; model; community; treatment; policy; regulation; infrastructure; literature; challenge; intervention

Introduction

Background

In the late 1970s a method was developed in the United States for the treatment of persons with severe mental disorders, the so-called Assertive Community Treatment (ACT) model [1]. An ACT team is a multidisciplinary team that includes case managers, psychiatrists, psychologists, and substance abuse specialists. The ACT model defines the following services to be provided by ACT teams: contact in the community and a holistic approach to care, for example, housing, medication, finances, everyday needs, and continuous coverage, which means that the ACT teams should be available for the patients at all times [1]. The ACT model estimates that the target group is 20% of the persons with a long-term severe mental disorder within a defined area [2]. Data concerning the patients and their pathways are usually displayed on a shared team whiteboard and discussed in daily team meetings. The purpose of the board is to maintain communication within the team, make sure no patients are left out, and make the daily meetings more efficient [3]. In later years, some ACT teams have used electronic

whiteboards and videoconferencing to communicate with patients that are able to use this kind of technology [4].

While the ACT model has shown great results in the United States, the results of some European implementations of the model were below expectations [5]. A justification for these results is the different health care models of each country and characteristics of the target group [5]. Additionally, in areas with low population density, the number of persons with severe mental disorders was too low to implement full ACT teams [2]. Because of these challenges, in the early 2000s, a variant of the ACT model intended for rural areas was implemented in the Netherlands, known as Flexible Assertive Community Treatment (FACT) [2]. The main difference between the 2 models is that the FACT model aims at supporting 100% of the persons with severe long-term mental disorders in the area [3], making it better suited for areas with low population density. This implies that many patients do not need continuous intensive follow-up, therefore, FACT teams provide individual case management [2]. As in the ACT model, patients at risk of relapse or readmission receive intensive follow-up from FACT teams. After the FACT model was implemented in the Netherlands, it has spread to a large number of health care teams in Sweden, Norway, and England, despite the lack of conclusive scientific evidence of its effectiveness [6].

FACT in Norway

In Norway, the government has the responsibility for specialist health care services, which are divided into 4 Regional Health Authorities. The Regional Health Authorities own the hospitals, which are organized as independent health trusts. Specialist mental health services can be provided by both the hospitals and community mental health centers. The responsibility for primary care and local services is assigned to the 356 municipalities. Most Norwegian FACT teams are organized as a cooperation between specialist mental health care and services from 1 or more municipalities [7]. Patients are referred to FACT teams by general practitioners or institutions in the specialist health care.

The ACT/FACT methods are recommended in the *National Health and Hospital Plan 2020-2023* [8] and *Coping With Life: The Government's Strategy for Good Mental Health (2017-2022)* [9]. In line with these recommendations, the Norwegian Directorate of Health has been granting funds for municipalities and health institutions to establish ACT/FACT teams since 2009 [8]. From 2009 to 2013, there were 14 ACT teams established in Norway. In 2013, the first FACT team was established, and in 2020, there were already approximately 70 FACT teams in Norway [10]. This shows the strong focus of the Norwegian Government in establishing FACT teams to support patients with severe mental disorders.

Norway has a population of 5,402,171 and an area of 323,808 km^2 [11]. Consequently, many persons have long travel distances to the nearest hospital or community mental health center. Therefore, FACT teams in rural areas of Norway typically support several municipalities, leading to distributed FACT teams situated on several locations.

In the context of this paper, we define eHealth as the use of information and communication technology (ICT) to improve efficiency, quality, and security in the health care delivery. The use of eHealth interventions as a solution for geographic and demographic challenges is one of the focus points of the Norwegian eHealth strategy [8]. In this paper, we discuss how the Norwegian eHealth infrastructure and regulations have affected the implementation of the FACT model in the country.

The overall aim of this paper is to conclude on how the implementation of the FACT model in Norway can be improved using eHealth interventions. To this end, we have examined the Norwegian policy regulating mental health services, eHealth regulations, eHealth infrastructure, relevant literature on FACT teams, and knowledge about the current implementation of FACT teams in Norway.

Methods

Norwegian Policy Regulating Mental Health Services

To get an overview of the governance of the Norwegian eHealth and mental health sectors we performed a content analysis of policy documents. We searched the governmental sites, The Norwegian government [12], The Norwegian Directorate of Health [13], and The Directorate of eHealth [14], for policy documents within the following main research topics: (1) eHealth, (2) health services delivery, and (3) health services delivery specific for mental health. The content analysis [15] identified the themes of data target by the policy documents. The content analysis included not only currently valid policy documents, but also documents that, even if they are no longer valid, still regulate mental health services.

Laws and Regulations Governing eHealth

Laws and regulations govern how health care workers do their work, how data are shared in the health care sector, and how ICT systems for health care can operate. For this reason, to understand implementation of ICT systems for FACT teams it is necessary to consider the relevant laws.

To identify relevant laws and regulations, we searched the Norwegian law database [16] for laws and regulations that are currently valid. We also studied the Norm for Information Security and Privacy in the Health and Care Sector (The Norm, from the Norwegian: *Normen for informasjonssikkerhet og personvern i helse- og omsorgstjenesten*), an industry norm that is developed and maintained by organizations and institutions in the Norwegian health care sector.

eHealth Infrastructure

The eHealth infrastructure has a large impact on what eHealth solutions are available, thus impacting how FACT teams operate. There are several national eHealth services in Norway that are facilitated by the national eHealth infrastructure.

In 2012, the Ministry of Health and Care Services published a governmental white paper—One Citizen–One Journal [17]—with the aim of presenting the goals for ICT development in health care in Norway. In this document the main strategy targeting the eHealth infrastructure indicated that electronic communication should be the way of communicating in the

Norwegian health care sector. The Norwegian Health Net (NHN) is an enterprise owned by the Ministry of Health and Care Services, which has the responsibility to manage, operate, and further develop the national eHealth infrastructure.

Among the services provided by the NHN the ones relevant for the FACT team are The Norwegian summary care record (SCR) and the Helsenorge portal [8,18]. The SCR is a collection of health information for patients that is available for all levels of health care in Norway. Helsenorge is a public portal for national digital health services in Norway. The portal contains health-related information for the citizens, personal health information, and various self-service solutions. Examples of self-service solutions are access to the SCR and information about prescriptions and vaccines. We studied official documents issued by the NHN to collect knowledge on important parts of the Norwegian eHealth infrastructure (ie, electronic messages, videoconferencing, and the SCR).

Relevant Literature on FACT Teams

To get an overview of research already done on eHealth interventions for ACT and FACT teams, we performed literature searches for original papers in the databases PsycINFO and PubMed. We searched PubMed for the large number of articles on eHealth research, and PsycINFO for additional articles with a focus on mental health.

To find relevant articles we made a list of keywords that describe ICT interventions. This list was combined with the search string "assertive community treatment" to find what related to ACT or FACT teams. The full search string is provided below:

> "assertive community treatment" AND (ehealth OR e-health OR telemedicine OR telepsychiatry OR ICT OR ehr OR digital OR technology OR video OR whiteboard)

We analyzed the titles and abstracts of the articles that resulted from the search string for inclusion according to the following predefined exclusion criteria: papers that did not report on ICT solutions for ACT or FACT teams, and papers that were not in English. We used the web tool Rayyan [19] for organizing inclusion decisions.

After applying exclusion criteria, we did a full-text analysis on the remaining papers.

FACT Implementation

The purpose of the Norwegian National Advisory Unit on Concurrent Substance Abuse and Mental Health Disorders (NKROP; from the Norwegian *Nasjonal kompetansetjeneste for samtidig rusmisbruk og psykisk lidelse*) is to run and support various projects and measures with the aim of enhancing health, quality of life, and functional level for persons with concurrent substance abuse and mental health disorders. One part of this purpose is supporting FACT teams with implementation of the FACT model. As part of the unit's role, NKROP makes available several documents regarding the implementation of the FACT model. We have studied the documents available at the unit's webpage to get an overview of guidelines for the FACT implementation in Norway. We selected the documents that provided information on the technical implementation of the FACT model.

The FACT Handbook [3] was written by one of the founders of the FACT model and was translated to Norwegian in 2013. This handbook describes the FACT model and how FACT teams work. We studied this document because it has been an important guideline in the practical implementation of FACT teams in Norway. Even though the handbook does not provide explicit information on the technical implementation of the FACT model, we included this document as it provides information that can inform the definition of requirements toward technology.

Norwegian FACT teams were evaluated in a report from 2020 [7]. We also studied this document because it shows many of the experiences of FACT teams.

A new published paper studied how Norwegian FACT teams are integrated into the service system [10]. This paper did not match our inclusion criteria for the literature search, because of a lack of focus on ICT solutions. However, it describes aspects important to the implementation of FACT teams, and thereby it was included in our study.

Results

Norwegian Policy Regulating Mental Health Services

The search of the governmental websites identified 6 policy documents relevant for the implementation of FACT teams. The documents were categorized under the main research topics as described below.

Main Research Topic 1: eHealth

The governmental white paper *One Citizen–One Journal* (in Norwegian: *Én innbygger – én journal*) [17], published in 2012 by the Ministry of Health and Care Services, aimed at presenting the goals for ICT development in health care in Norway. The main goal of this white paper was to ensure health care workers have easy and secure access to patient and user information; citizens have access to easy and simple digital solutions; and data are available for quality improvement, monitoring, management, and research. The paper also stated that the goal of the government is that all written communication in health care should be electronic. To reach these goals the government wanted to modernize the ICT platform, and work toward a national ICT solution for the whole health and care sector. The white paper also points out challenges of a lack of integration between systems.

Main Research Topic 2: Health Services Delivery

The Coordination Reform (in Norwegian: *Samhandlingsreformen*), published in 2009 [20], was a report that described a reform of Norwegian health services. The report pointed out 3 primary challenges: services that are poorly coordinated, lack of focus on disease prevention, and an increase in the prevalence of chronic diseases due to an aging population. To meet these challenges, 5 measures were proposed: (1) There should be a stronger focus on user involvement and better patient pathways; (2) Municipalities should have an increased focus on prevention, early intervention, treatment, and follow-up; (3)

Economic incentives should be put in place, where municipalities cofinance the specialist health care, and are economically responsible for patients who are ready to be discharged from the hospital; (4) Specialist care should be more focused on specialist tasks. This should be achieved by a better division of work between primary and specialist care. Also, a stronger focus on patient pathways would make patients in need of specialist care get the right treatment; (5) The government should also have a more holistic view on the health care and the patients when prioritizing needs. The report also pointed out that more integrated patient pathways would require new ways of using ICT solutions and emphasized the challenges of ICT systems not communicating well enough. Also, many beds in specialist mental health are used by patients who should have received stronger follow-up from municipalities instead. ACT teams are described as one way of achieving this.

The *National Health and Hospital Plan 2016–2019* (in Norwegian: *Nasjonal helse- og sykehusplan 2016–2019*) [21] presented the governments goals for the development of specialist health care for the period 2016-2019. The overall goals were to focus on patient-centered care; prioritize the field of mental health and addiction; renew, improve, and simplify services; contribute to enough health care workers with the right competences; improve quality and patient safety; improve division of responsibilities and cooperation between hospitals; and improve emergency medicine outside the hospitals. The plan states that the field of mental health and addiction should be prioritized, and that specialist health care should be provided close to where the patient lives. Specialist health care should cooperate with municipal services, and for the patient, the health care service should appear integrated. The plan also stated explicitly that large cities/towns should have ACT teams. ICT systems should also support good work processes and patient pathways.

The *National Health and Hospital Plan 2020-2023* (in Norwegian: *Nasjonal helse- og sykehusplan 2020-2023*) [8] was a revision of the *National Health and Hospital Plan 2016-2019*. The main goal of the new plan was to achieve a patient-centered care system in a sustainable way. The focus areas of the plan were better cooperation between specialist care and municipal services, improvement of mental care, focus on technology, improved digitalization, and to ensure health care workers with the right competence are available. It also underlined the need for cooperation between specialist health care and municipal health care. To improve patient pathways, the plan describes a need for ICT systems to share information between the different levels of health care. According to the plan, some of the most important measures to improve digitalization of the health care are to continue modernizing electronic health records (EHRs), improving access and availability of health information, and supporting digitalization of health and care services in the municipalities. The plan also promotes team-based methods of working, such as ACT and FACT.

Main Research Topic 3: Health Services Delivery Specific for Mental Health

Overview

The *Norwegian National Action Plan in Mental Health (1999-2008)* (in Norwegian: *Opptrappingsplan for psykisk helse 1999-2008*) [22] indicated several issues with the mental health care services. This included lack of prevention, poor services in the municipalities, often too short inpatient stays, and a lack of follow-up after discharge from a hospital. In this regard, the main goal of the plan was to strengthen mental health care and make more holistic and coherent care services available. It also pointed out that coordinated services were needed for persons with serious mental health services.

Coping With Life: The Government's Strategy for Good Mental Health (2017-2022) (in Norwegian: *Mestre hele livet - Regjeringens strategi for god psykisk helse 2017-2022*) [9] was the first mental health strategy to follow The Norwegian National Action Plan in Mental Health (1999-2008). This document presented a holistic strategy for the field of mental health. The aims were to make mental health a part of public health work; promote inclusion into the society; focus on patient-centered care; improve knowledge, quality, research, and innovation in the services; and have a special focus on children and youth. The white paper also states that there is a need for research-based development of digital tools and guided internet-assisted treatment, and that experiences with ACT and FACT teams show that they lead to reductions in involuntary treatment. The paper also focuses on labor participation and housing, areas that are also important for FACT teams.

Based on the content identified from the policy documents that were deemed relevant, we defined the following themes of data: (1) eHealth solutions and infrastructure, (2) organizational management, and (3) health care services coordination.

The focus of the content included in each theme of data was the basis to further identify subthemes of data. The relevant content from the policy documents was then categorized under the subthemes of data, as presented below.

eHealth Solutions and Infrastructure

Electronic communication refers to ICT communication systems between health care institutions for the exchange of patient information [17].

National digital services and ICT infrastructure include digital services for citizens, process support that should be facilitated by the systems in use, and ICT infrastructure that should support the national deployment of eHealth [8,17,22].

Data sharing and access concern interoperability of systems in the sense of sharing of patient information between different levels and institutions in health care, and integration of ICT systems [8,17,20].

eHealth records refer to the requirement to update EHR systems to comply with current user needs [8].

Secondary use of data means the reuse of information for quality improvement, monitoring, management, and research [17].

Organizational Management

Process support identifies the need for ICT to support care pathways [21].

Access to care refers to health care services being provided to patients when and where needed [20-22].

Care service delivery concerns how patients receive health care services, including national patient pathways [8,20-22].

Financing and prioritization of mental health care refer to the national strategy for the field of mental health [20-22].

Competence refers to ensuring health care workers with the right competence are available [8,21].

Prevention, early intervention, and follow-up refer to municipalities taking responsibility for preventive care, early intervention, and follow-up [22].

Health Care Services Coordination

Coordinated services refer to the cooperation between different levels of health care, including health ICT development [8,21,22].

Mental health services indicate that mental health services should be coordinated so that patients receive care services from the appropriate level of care [9,20].

Team-based health care services refer to the use of team-based methods in health care, such as ACT and FACT [9].

The "eHealth solutions and infrastructure" theme is discussed in 4 policy documents, the "Organizational management" theme is also discussed in 4 policy documents, and the "Health care services coordination" theme is discussed in 5 documents.

Laws and Regulations Governing eHealth

There are several Norwegian laws and regulations governing eHealth that are relevant for FACT teams. These include the Personal Data Act [23], the Patient Journal Law [24], the Health Personnel Law [25], and the Patient journal regulation [26].

The Personal Data Act applies the requirements of European Union's General Data Protection Regulation (GDPR). The Personal Data Act and GDPR state that a data controller is the person or institution responsible for ensuring that the data are treated according to the principles relating to personal data processing. The data controller must establish the necessary technical and organizational measures to ensure that the laws are followed [27]. A data processor is a person or institution that processes data on behalf of a data controller [28]. Medical research on humans, human biological material, or health information needs to be preapproved by the Regional Committees for Medical and Health Research Ethics for the relevant region [29]. The committee does an appraisal of research ethics and if the project is in accordance with relevant laws.

The Patient Journal Law §19 allows for relevant and necessary information about a patient to be made available for health care workers when it is needed for providing, administering, or ensuring the quality of health care [24]. This is regardless of where the patient was treated earlier and how the health service is organized [21].

The Health Personnel Law §25 states that unless the patient refuses, cooperating health care workers can be given access to patient information when this is necessary to provide health care [25]. For 2 or more health care institutions to both have access and be able to update common EHR information, the Patient Journal Law §9 states that there needs to be a written agreement about how the institutions shall cooperate [24]. The Patient Journal Regulation [26] states that EHR information should only be available for health care workers who can confirm their identity in a secure manner.

The Norm gives institutions that follow it the necessary technical and organizational tools to ensure security and privacy when processing health information. The Norm includes several fact sheets and guides for processing health information. Some fact sheets that are relevant for FACT teams are for electronic messages, internal communication, and access to health information between organizations. A fact sheet from The Norm [30] specifies that to ensure security, data should be encrypted, authentication is necessary to access data, and access should be logged and monitored. To request sharing of data between institutions, a risk assessment is required for, and the partners who share information need to have an agreement in place about the data sharing. Patients have the right to see information about who has had access to their patient journal.

In practice, EHR data are usually easily accessible for health care workers in the institution that is responsible for the data, but harder to access for health care workers outside the institution. To circumvent this, health care workers who need access to patient information from an institution they are not affiliated to have sometimes been hired in a so-called zero percent position or simplified employment in the institution. This gives them access to the institution's EHR, as employees of the institution.

eHealth Infrastructure

Situation in Norway

Norway is divided into 4 health regions that are run by state-owned regional health authorities [31]. The regional health authorities are responsible for offering specialist health care to the population in the region. Each year, the Ministry of Health and Care Services gives the regional health authorities a commissioner's document, which stands as a reference document concerning the needs for the health care sector in each region. How these needs are implemented are defined by each regional health authority. This means that the different regional health authorities have different plans for what eHealth solutions they will develop and use.

EHR Systems

The hospitals in 3 of the 4 health care regions are using the EHR named after the provider DIPS AS. The exception is the hospitals in the Central Norway Regional Health Authority, which are using DocuLive provided by Siemens AS. In the Norwegian primary care sector, there are several different EHR system vendors. However, the ongoing project Helseplattformen

[32] is working on the implementation of a common EHR system for specialist and primary care in the central region of Norway.

For Norwegian FACT, which is organized as a cooperation between specialist mental health care and services from 1 or more municipalities, this means that the team members have different EHR systems available. Because of this, the teams must do the extra work of documenting in several EHRs or accept that some of the EHR systems will lack data.

Electronic Messages

In the Norwegian health care sector, standardized electronic messages are used to communicate between the different levels of care. Some of the types of messages sent are referrals, discharge letters, blood test results, and messages regarding sick leaves.

To ensure correct addressing for electronic and non-electronic communication in health care, there is a national address register for health care institutions. This allows for the standardized electronic messages to be sent from the sender's EHR to the receiver through the Health Net [33]. Specialist care and municipal mental health services are among the partners who can use standardized electronic messaging. An evaluation of Norwegian FACT teams [7] showed that there are mixed experiences with standardized electronic messaging when used in the context of FACT teams. The report identified discharge letters from hospitals and documentation of medication as standardized electronic messages that were useful. By contrast, the report stated that messages from the outpatient clinic were too slow and identified this as a barrier for cooperation.

Videoconference

Several solutions have been used for video consultations with patients in Norway. The Norwegian national health portal Helsenorge [18] offers seamless integration of third-party videoconferencing solutions from their portal. Several different vendors have been approved for video integration. This means that there is no common solution for the use of videoconferencing for patient consultations in Norway. Even though FACT teams are moving toward the use of the video service provided by Helsenorge [18], at present, each FACT team can decide on how to implement and what videoconferencing solution to use following the guidelines of the health region they are in.

The Norwegian Summary Care Record

The SCR is a collection of health information for patients, which is available for all levels of health care in Norway. The goal of the SCR is to provide health care workers in different institutions quick access to important health information about patients. Patients have the possibility to choose not to have an SCR, and they can limit who has access and can deny access to parts of their SCR. The SCR can contain medication list, contact information, admission history, and critical information including a psychiatric emergency plan. Citizens can register information about their primary contact person, disease history, special needs, and information about being an organ donor. Somatic and mental health are treated equivalent in the SCR, and the admission list may also show visits to the psychiatric ward [34].

FACT team workers have access to the SCR like other health care workers and can use it to access medication lists and other relevant information.

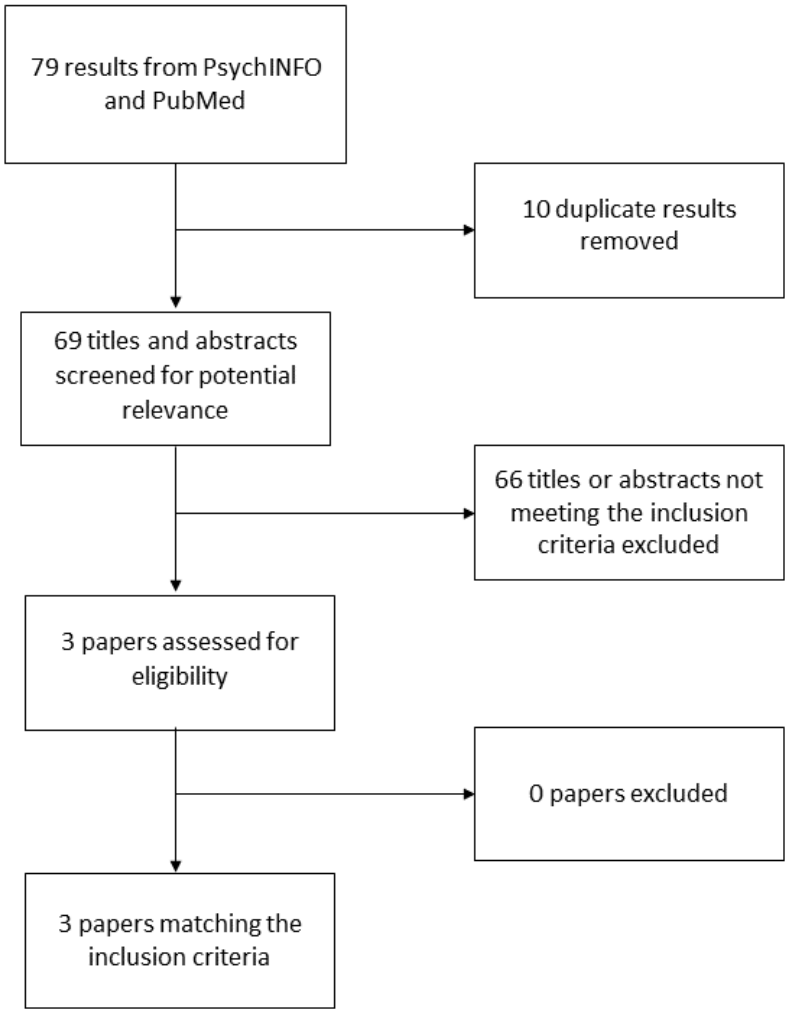
Relevant Literature on FACT Teams

Our searches returned 59 results in PsycINFO and 20 results in PubMed. We removed 10 duplicates, for a total of 69 results. After a review of the titles, 18 articles were selected for an abstract review, and 51 was excluded. In the abstract review, 3 articles were selected for a full-text review. These 3 articles matched our inclusion criteria. The 3 included articles describe eHealth interventions targeting ACT patients.

Figure 1 presents the results based on the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) guidelines [35].

Ben-Zeev et al [36] described a randomized control trial on texting mobile intervention added to ACT. In the trial, mental health workers had recovery-oriented texting exchanges based on the patients' individual needs. The trial was feasible, acceptable, and safe. Looijmans et al [37] described a study protocol for a randomized controlled trial of a web tool intervention designed to improve cardiometabolic health in patients with severe mental disorders, including ACT patients. The intervention group received motivational interviews and a web tool that covers behavior awareness, lifestyle knowledge, motivation, and goal setting. Swanson and Trestman [4] described the use of a videoconference solution to supplement the face-to-face relationship with the team's psychiatrist. The videoconference solution is used for crisis intervention and for augmentation of the established face-to-face treatment. The solution has been accepted by the staff and the patients, and seriously ill patients have been able to use the solution. The solution has also reduced travel time for the psychiatrists.

Figure 1. PRISMA flowchart of the literature search and article selection.



FACT Implementation

The FACT Handbook [3] is a description of the FACT model written by one of the founders of the model. It was commissioned by the Norwegian Directorate of Health and has been used as a guide for implementing new FACT teams in Norway.

A report on the evaluation of Norwegian FACT teams [7] showed that all 7 teams evaluated used electronic whiteboards to display patient information. Some teams had bought a commercial solution for a whiteboard, while others had developed them locally or used the solutions from other teams. The report also demonstrated that there are different approaches to using EHRs in Norwegian FACT teams. Two of the 7 teams document in 2 separate EHRs: 1 for primary care and 1 for specialist care. However, 5 teams use the EHR only for specialist care. In these FACT teams, team members who work in primary care are given "zero percent positions" in the specialist care. This means that they are formally hired, to allow access to the specialist health care EHR. Interviews conducted with the FACT teams' cooperating partners also emphasize the challenges of working with different EHR systems. Some cooperating partners found use of standard electronic messages useful for improving coordination. Others found that the sent standard electronic messages took too long before they were received, and that this did hinder the ability to provide integrated services to the patients. The report also pointed out that different regulations governing FACT teams and municipalities and the multiple EHR systems used by the teams hindered service integration. One of the main recommendations of the report is that digital communication solutions should be facilitated to improve coordination.

A study on how Norwegian FACT teams are integrated into the Norwegian health care system [10] showed that FACT teams reduce complexity and reassure other services by taking responsibilities for treatment and follow-up of patients. However, the study also showed that there is a lack of common communication systems, making exchange of information harder and more time consuming. The study also reported that not all patient information is available in the different EHR systems in use.

Discussion

Norwegian Policy Regulating Mental Health Services

Policy documents identify ACT and FACT teams as a wanted part of the Norwegian health care system. ACT and FACT have been recommended in the policy documents *Coping With Life: The Government's Strategy for Good Mental Health (2017-2022)* [9], the *National Health and Hospital Plan*

2016-2019 [21], and the *National Health and Hospital Plan 2020-2023* [8].

Considering the number of policy documents targeting each theme of data, we conclude that the 3 themes of data considered are relevant for the implementation of FACT teams in Norway.

eHealth Solutions and Infrastructure

Policy documents state that better eHealth solutions are needed in health care. Examples of this are the *One Citizen–One Journal* [17], which stated that the goal of the government is to make electronic communication as the standard way of written communication within health care; and the *National Health and Hospital Plan 2016-2019* [21], which states that ICT systems should also support good work processes and patient pathways. The *National Health and Hospital Plan 2020-2023* [8] renews this vision by stating that ICT systems are needed to share information between the different levels of health care. The plan also states that further modernization of EHRs is needed.

While the policy documents do not specifically discuss ICT challenges for FACT teams, these issues are also present in FACT teams, with the lack of horizontal and vertical collaboration in health care. This is mainly seen in the lack of electronic communication and sharing of patient information between the team members. The needs from FACT teams are the same as those described in the policy documents: National digital services and ICT infrastructure. The implementation of the work processes defined by the FACT model is not supported by the existing eHealth system. For them to work as expected, there should be similar implementation of the FACT model across the country. Besides, the ICT infrastructure needs to evolve to comply with the requirements of the FACT model, which implies the deployment of national digital services for FACT teams. Secondary use of health data relevant for FACT teams, such as quality improvement, monitoring, management, and research, will also be hindered until better eHealth systems are in place.

Organizational Management

One vital aspect of the policy documents is describing the organization of the health care service. Three of the identified strategy documents [20-22] state that the field of mental health should be prioritized. Four documents also state that ACT or FACT teams should be implemented [8,9,20,21]. Four of the 5 identified strategy documents refer to access to care and care service delivery, with an emphasis on implementing standardized patient pathways [8,20-22]. In this context, 1 policy document points to the lack of follow-up after hospital discharge as an issue [22]. One of the primary goals of FACT teams is to improve follow-up for their patient group.

The essence of the FACT model is the delivery of integrated care through multidisciplinary teams. As emphasized in the policy documents, FACT teams also need to have access to the required competence to be able to provide integrated care services with the expected quality. In practice, FACT teams must make work-arounds to deliver integrated care services as the existing ICT infrastructure does not facilitate process support, as mentioned in the policy documents.

Health Care Services Coordination

Even though team-based methods of care delivery, such as ACT and FACT, are promoted [8], the lack of coordination among the Norwegian health care services is one of the main challenges of the sector at present. This has been described in general for the health care sector in the *Coordination Reform* [20], as well as in the policy documents specific for mental health: *The Norwegian National Action Plan in Mental Health (1999-2008)* [22] and *Coping With Life: The Government's Strategy for Good Mental Health (2017-2022)* [9]. The lack of coordination between the different levels of care and between institutions greatly affects the way FACT teams are implemented, as the cornerstone of the FACT model is the delivery of coordinated services.

FACT teams in Norway have found ways to overcome the lack of coordination in the health care sector to still be able to deliver coordinated services. This includes the aforementioned "zero percent positions," which allow all team members access to EHR information from specialist care. The evidence that FACT teams provide better coordinated services and contribute to the reduction of both emergency admissions and forced admissions has resulted in the implementation of several FACT teams in Norway [7].

eHealth Regulations

One barrier for FACT teams is eHealth regulations. However, the Patient Journal Law §19 allows necessary information about a patient to be made available for health care workers when it is needed for providing, administrating, or ensuring the quality of health care. Besides, the Health Personnel Law §25 states that health information can be given to cooperating health care workers when this is necessary to provide health care. However, in practice it is often hard for health care workers to access EHR data from institution where they do not work. This shows that current ICT solutions do not take advantage of what is made possible by the laws and there is a need for systems to allow EHR access to relevant data, while preserving the privacy of the information.

In the context of the personal data law and GDPR, hospitals or health trusts take the role of data controllers, responsible for EHR data in their areas. EHR vendors are data processors that process data on behalf of the controllers. There is a need for an agreement between the data controller and data processor, stating the obligations of each of the partners.

In addition to the laws, the Norm is a helpful guideline when implementing ICT solutions in Norway, which provides necessary information for practical implementation. The Norm also specifies responsibilities of data controllers and data processors.

eHealth Infrastructure

Lack of integration and coordination between services and technologies has been seen as an issue for use of eHealth within psychiatry [38]. Also in Norway, implementation of FACT teams has been affected by the eHealth infrastructure. Sharing of data has been reported as a problem for Norwegian FACT teams [7]. The specialist and primary care in Norway use

different EHR systems, and the technical infrastructure in place does not allow easy exchange of information between systems. This leads to challenges for FACT teams that often have members from both specialist and primary health care.

Different health regions in Norway can also have different implementation of regulations. There are various ways of bypassing these challenges, but ideally the EHR systems should be able to display relevant EHR data to the health care workers who need the information for treating their patients, even when the information is stored in different systems. In general, there are challenges with sharing information between the different levels of health care and different regions on the same level. Use of electronic messages is one well-established way of doing this and is useful in many contexts. However, the use is still limited to standard messages, such as referrals and discharge letters.

Videoconferencing has been used for both communication between health care workers and communicating with patients. The COVID-19 pandemic increased its use in many FACT teams. Video consultations with patients might be a useful tool for FACT teams, but it might not be suited to all their patients. Various solutions for video consultations are used in Norway, and some of these have integration to the portal Helsenorge [18] and the calendar in DIPS.

The only data that are shared with all health care workers in Norway are the SCR. This record includes information on medication, which can be useful for FACT teams. However, the intention of the SCR is not to support clinical cooperation.

Relevant Literature on FACT Teams

Our literature search returned 3 articles that matched our inclusion criteria, which reported on the use of an SMS text message intervention [36], a web tool [37], and the use of videoconferencing [4]. These 3 articles showed that eHealth interventions can successfully target ACT patients. However, we found no articles discussing ICT challenges for ACT or FACT teams themselves. While we did not perform a full literature review, this implies that there is little or no research on this topic. Health care is organized differently across countries, and many of the challenges regarding eHealth infrastructure and regulations described in this article are not relevant in other countries. Thus, the lack of articles on this topic is not surprising.

FACT Implementation

The evaluation of FACT teams has shown their positive effects on patients, including reduction of inpatient days, with a larger reduction in compulsory inpatient days [7]. The evaluation also highlights some of the ICT challenges the teams face, including EHR access. A study of how FACT teams fit into the Norwegian service model also showed that the cooperating partners of FACT teams think the lack of common communication systems and EHR systems is a challenge [10]. These findings show that, despite some issues with EHR systems, the implementation of FACT teams in Norway has been successful and can be expected to continue.

Conclusions

Weaknesses in the Norwegian eHealth infrastructure have been a barrier for an easy implementation of FACT teams in Norway. The FACT evaluation report identifies the sharing of information between the different levels of health care as a main shortcoming of the existing eHealth infrastructure. FACT teams need eHealth systems that allow easy and secure sharing of health information with the team members and other relevant health care workers to provide better care. There is also a lack of research on the ICT challenges facing FACT teams. This means that there is a need for research studying the eHealth challenges and needs of FACT teams in greater detail. Furthermore, there is a need to explore how eHealth solutions should be designed to support FACT teams in a Norwegian context. This is something we will focus on in future work.

Acknowledgments

This project is funded by The Research Council of Norway (project number 288722). We thank our project partners Norwegian National Center for Concurrent Substance Abuse and Mental Disorders, Innlandet Hospital Trust, National Resource Center for Community Mental Health (NAPHA), Norwegian Centre for e-Health Research, University of Tromsø, and Inn University of Applied Sciences. We also thank our colleague Karianne Lind for assisting in the literature search.

Conflicts of Interest

None declared.

References

1. Bond GR, Drake RE. The critical ingredients of assertive community treatment. World Psychiatry 2015 Jun;14(2):240-242 [FREE Full text] [doi: 10.1002/wps.20234] [Medline: 26043344]
2. van Veldhuizen JR. FACT: a Dutch version of ACT. Community Ment Health J 2007 Aug;43(4):421-433. [doi: 10.1007/s10597-007-9089-4] [Medline: 17514502]
3. van Veldhuizen JR. FACT - Visjon, modell og organisering av FACT-modellen [FACT - Vision, model and organization of the FACT model]. Nasjonal kompetansetjeneste for samtidig rusmisbruk og psykisk lidelse. 2013. URL: https://rop.no/globalassets/dokumenter/fact-handbok.pdf [accessed 2021-12-29]
4. Swanson CL, Trestman RL. Rural assertive community treatment and telepsychiatry. J Psychiatr Pract 2018 Jul;24(4):269-273. [doi: 10.1097/PRA.0000000000000313] [Medline: 30427810]

5. Nugter MA, Engelsbel F, Bähler M, Keet R, van Veldhuizen R. Outcomes of FLEXIBLE Assertive Community Treatment (FACT) Implementation: a prospective real life study. Community Ment Health J 2016 Nov;52(8):898-907 [FREE Full text] [doi: 10.1007/s10597-015-9831-2] [Medline: 25648552]
6. Svensson B, Hansson L, Markström U, Lexén A. What matters when implementing Flexible Assertive Community Treatment in a Swedish healthcare context: A two-year implementation study. International Journal of Mental Health 2017 Jul 20;46(4):284-298. [doi: 10.1080/00207411.2017.1345041]
7. Landheim A, Odden S. Evaluering av FACT-team i Norge - Sluttrapport [Evaluation of FACT teams in Norway]. 2020. URL: https://tinyurl.com/2p99bxhb [accessed 2021-12-29]
8. Ministry of Health and Care Services. Nasjonal helse- og sykehusplan 2020-2023. 2019. URL: https://www.regjeringen.no/no/dokumenter/nasjonal-helse--og-sykehusplan-2020-2023/id2679013/ [accessed 2021-12-29]
9. Departementene. Mestre hele livet - Regjeringens strategi for god psykisk helse (2017-2022). 2017. URL: https://www.regjeringen.no/no/dokumenter/mestre-hele-livet/id2568354/ [accessed 2021-12-29]
10. Trane K, Aasbrenn K, Rønningen M, Odden S, Lexén A, Landheim A. Flexible assertive community treatment teams can change complex and fragmented service systems: experiences of service providers. Int J Ment Health Syst 2021 Apr 23;15(1):38 [FREE Full text] [doi: 10.1186/s13033-021-00463-1] [Medline: 33892769]
11. Statistics Norway. Population. Statistics Norway. 2021. URL: https://www.ssb.no/en/befolkning/folketall/statistikk/befolkning [accessed 2021-12-29]
12. The Norwegian Government. 2021. URL: https://www.regjeringen.no/no/id4/ [accessed 2021-12-30]
13. The Norwegian Directorate of Health. 2021. URL: http://www.helsedirektoratet.no/ [accessed 2021-12-30]
14. Direktoratet for e-helse [The Directorate of eHealth]. 2021. URL: http://www.ehelse.no/ [accessed 2021-12-30]
15. Bowen GA. Document analysis as a qualitative research method. Qualitative Research Journal 2009 Aug 03;9(2):27-40. [doi: 10.3316/QRJ0902027]
16. Lovdata. 2021. URL: https://lovdata.no/ [accessed 2021-12-30]
17. Ministry of Health and Care Services. Én innbygger – én journal. 2012. URL: https://www.regjeringen.no/no/dokumenter/meld-st-9-20122013/id708609/ [accessed 2021-12-20]
18. Helsenorge. 2021. URL: https://www.helsenorge.no/ [accessed 2021-12-30]
19. Rayyan. 2021. URL: https://www.rayyan.ai/ [accessed 2021-12-30]
20. Ministry of Health and Care Services. Samhandlingsreformen [Coordination Reform]. 2009. URL: https://www.regjeringen.no/contentassets/d4f0e16ad32e4bbd8d8ab5c21445a5dc/no/pdfs/stm200820090047000dddpdfs.pdf [accessed 2021-12-29]
21. Ministry of Health and Care Services. Nasjonal helse- og sykehusplan (2016-2019) [National Health and Hospital Plan 2016-2019]. 2015. URL: https://www.regjeringen.no/no/dokumenter/meld.-st.-11-20152016/id2462047/ [accessed 2021-12-29]
22. Ministry of Health and Care Services. Opptrappingsplan for psykisk helse 1999 - 2008 [The Norwegian National Action Plan in Mental Health (1999-2008)]. 1998. URL: https://www.regjeringen.no/no/dokumenter/stprp-nr-63-1997-98-/id201915/ [accessed 2021-12-29]
23. Ministry of Justice and Public Security. Lov om behandling av personopplysninger (personopplysningsloven) [The Personal Data Act]. 2018. URL: https://lovdata.no/dokument/NL/lov/2018-06-15-38 [accessed 2021-12-29]
24. Ministry of Health and Care Services. Lov om behandling av helseopplysninger ved ytelse av helsehjelp (pasientjournalloven) [The Patient Journal Law]. 2021. URL: https://lovdata.no/dokument/NL/lov/2014-06-20-42 [accessed 2021-12-29]
25. Ministry of Health and Care Services. Lov om helsepersonell m.v. (helsepersonelloven) [The Health Personnel Law]. 2021. URL: https://lovdata.no/dokument/NL/lov/1999-07-02-64 [accessed 2021-12-29]
26. Ministry of Health and Care Services. Forskrift om pasientjournal (pasientjournalforskriften) [The Patient Journal Regulation]. 2021. URL: https://lovdata.no/dokument/SF/forskrift/2019-03-01-168 [accessed 2021-12-29]
27. Data Protection Authority. Hva er en behandlingsansvarlig?. 2019. URL: https://www.datatilsynet.no/rettigheter-og-plikter/virksomhetenes-plikter/databehandleravtale/behandlingsansvarlig-og-databehandler/hva-er-en-behandlingsansvarlig/ [accessed 2021-12-29]
28. Data Protection Authority. Hva er en databehandler? Datatilsynet.no.: Data Protection Authority; 2019. URL: https://tinyurl.com/bdc9p6an [accessed 2021-12-29]
29. Ministry of Health and Care Services. Lov om medisinsk og helsefaglig forskning (helseforskningsloven). Lovdata.no.: Ministry of Health and Care Services; 2021. URL: https://lovdata.no/dokument/NL/lov/2008-06-20-44 [accessed 2021-12-29]
30. The Directorate of eHealth. Sikkerhets- og samhandlingsarkitektur ved tilgang til helseopplysninger mellom virksomheter. ehelse.no.: The Directorate of eHealth; 2018. URL: https://www.ehelse.no/normen/faktaark/faktaark-20c-sikkerhets-og-samhandlingsarkitektur-ved-tilgang-til-helseopplysninger-mellom-virksomheter [accessed 2021-12-29]
31. Ministry of Health and Care Services. Oppdragsdokument. 2021. URL: https://www.regjeringen.no/no/tema/helse-og-omsorg/sykehus/styringsdokumenter1/oppdragsdokument/id535564/ [accessed 2021-12-29]
32. Helseplattformen. 2021. URL: https://helseplattformen.no/ [accessed 2021-12-29]
33. Norsk Helsenett. Veileder for elektronisk meldingsutveksling. 2017. URL: https://www.nhn.no/samhandlingsplattform/veileder-for-elektronisk-meldingsutveksling [accessed 2021-12-29]

34. Jensen BA. Alt eller ingenting. TNPF 2017;55(11):1105-1106. [doi: 10.52734/dk6ortsn]
35. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. BMJ 2009 Jul 21;339:b2535 [FREE Full text] [doi: 10.1136/bmj.b2535] [Medline: 19622551]
36. Ben-Zeev D, Buck B, Meller S, Hudenko WJ, Hallgren KA. Augmenting evidence-based care with a texting mobile interventionist: a pilot randomized controlled trial. Psychiatr Serv 2020 Dec 01;71(12):1218-1224 [FREE Full text] [doi: 10.1176/appi.ps.202000239] [Medline: 32631130]
37. Looijmans A, Jörg F, Bruggeman R, Schoevers R, Corpeleijn E. Design of the Lifestyle Interventions for severe mentally ill Outpatients in the Netherlands (LION) trial; a cluster randomised controlled study of a multidimensional web tool intervention to improve cardiometabolic health in patients with severe mental illness. BMC Psychiatry 2017 Mar 21;17(1):107 [FREE Full text] [doi: 10.1186/s12888-017-1265-7] [Medline: 28327086]
38. Nicholas J, Huckvale K, Larsen ME, Basu A, Batterham PJ, Shaw F, et al. Issues for eHealth in psychiatry: results of an expert survey. J Med Internet Res 2017 Feb 28;19(2):e55 [FREE Full text] [doi: 10.2196/jmir.6957] [Medline: 28246068]

Abbreviations

ACT: Assertive Community Treatment
EHR: electronic health record
FACT: Flexible Assertive Community Treatment
GDPR: General Data Protection Regulation
ICT: information and communications technology
NHN: Norwegian Health Net
PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses
SCR: summary care record

Edited by G Eysenbach; submitted 19.07.21; peer-reviewed by P Pednekar, SK Mukhiya; comments to author 30.07.21; revised version received 01.10.21; accepted 27.11.21; published 10.01.22.

Please cite as:
Bønes E, Granja C, Solvoll T
Implementation of the Flexible Assertive Community Treatment (FACT) Model in Norway: eHealth Assessment Study
J Med Internet Res 2022;24(1):e32220
URL: https://www.jmir.org/2022/1/e32220
doi:10.2196/32220
PMID:35006087

Original Paper

Perspectives of Policy Makers and Service Users Concerning the Implementation of eHealth in Sweden: Interview Study

Margit Neher[1], PhD; Annette Nygårdh[2], PhD; Anders Broström[2,3], PhD; Johan Lundgren[4], PhD; Peter Johansson[4], PhD

[1]Department of Rehabilitation, School of Health and Welfare, Jönköping University, Jönköping, Sweden

[2]Department of Nursing Sciences, School of Health and Welfare, Jönköping University, Jönköping, Sweden

[3]Department of Clinical Neurophysiology, Linköping University Hospital, Linköping, Sweden

[4]Department of Health, Medicine and Caring Sciences, Linköping University, Linköping, Sweden

Corresponding Author:
Margit Neher, PhD
Department of Rehabilitation
School of Health and Welfare
Jönköping University
Gjuterigatan 5
Jönköping, 551 11
Sweden
Phone: 46 70 7892336
Email: margit.neher@ju.se

Abstract

Background: Increasing life spans of populations and a growing demand for more advanced care make effective and cost-efficient provision of health care necessary. eHealth technology is often proposed, although research on barriers to and facilitators of the implementation of eHealth technology is still scarce and fragmented.

Objective: The aim of this study is to explore the perceptions concerning barriers to and facilitators of the implementation of eHealth among policy makers and service users and explore the ways in which their perceptions converge and differ.

Methods: This study used interview data from policy makers at different levels of health care (n=7) and service users enrolled in eHealth interventions (n=25). The analysis included separate qualitative content analyses for the 2 groups and then a second qualitative content analysis to explore differences and commonalities.

Results: Implementation barriers perceived by policy makers were that not all service users benefit from eHealth and that there is uncertainty about the impact of eHealth on the work of health care professionals. Policy makers also perceived political decision-making as complex; this included problems related to provision of technical infrastructure and lack of extra resources for health care digitalization. Facilitators were policy makers' conviction that eHealth is what citizens want, their belief in eHealth solutions as beneficial for health care practice, and their belief in the importance of health care digitalization. Barriers for service users comprised capability limitations and varied preferences of service users and a mismatch of technology with user needs, lack of data protection, and their perception of eHealth as being more time consuming. Facilitators for service users were eHealth technology design and match with their skill set, personal feedback and staff support, a sense of privacy, a credible sender, and flexible use of time.There were several commonalities between the 2 stakeholder groups. Facilitators for both groups were the strong impetus toward technology adoption in society and expectations of time flexibility. Both groups perceived barriers in the difficulties of tailoring eHealth, and both groups expressed uncertainty about the care burden distribution. There were also differences: policy makers perceived that their decision-making was very complex and that resources for implementation were limited. Service users highlighted their need to feel that their digital data were protected and that they needed to trust the eHealth *sender*.

Conclusions: Perceptions about barriers to and facilitators of eHealth implementation varied among stakeholders in different parts of the health care system. The study points to the need to reach an enhanced mutual understanding of priorities and overcome challenges at both the micro and macro levels of the health care system. More well-balanced decisions at the policy-maker level may lead to more effective and sustainable development and future implementation of eHealth.

(J Med Internet Res 2022;24(1):e28870) doi:10.2196/28870

KEYWORDS

clients; computer-assisted therapy; consultation telehealth; decision-makers; implementation; patients; politicians; qualitative methods; remote; mobile phone

Introduction

Background

Health care systems around the world are struggling to satisfy the health care needs of their populations. Because of the increasing population life span and the growing demand for more advanced care, the need for effective and cost-efficient provision of health care is a matter of national concern for many nations [1]. eHealth technology is often proposed as a promising way forward for better and more cost-efficient health care solutions. This *umbrella term*, which includes many different digital health care solutions such as telemedicine and the use of mobile messaging devices, has become a buzzword in many countries and is defined in the World Health Assembly resolution WHA58.28 (2005) as "the cost-effective and secure use of ICT in support of health and health-related fields" [2]. Many European health care systems see a proliferation in eHealth technologies, and policy decisions to implement them seem to be made despite the scarcity of convincing research data concerning both claims of cost-effectiveness and their real impact on the quality and safety of health care [3,4]. For example, in Sweden, the implementation of eHealth technologies has been identified as an important strategy toward the future quality provision of population health care. eHealth technology in different forms has been introduced in many different Swedish health care settings, with the hope of better quality of care, better patient empowerment in care, and lower health care costs [5]. At the national level, ambitious political goals have been set to become world leaders in using the opportunities offered by eHealth by the year 2025 [5].

Research on barriers to and facilitators of the implementation of eHealth technology is still scarce and fragmented; however, some reviews have collated what is currently known [6,7]. Implementation was found to have a multilevel complexity: the technology, individual health care providers, inner and outer settings, and process of implementation [7]. Greenhalgh et al [8] constructed an empirically based framework that described implementation complexity in 7 domains: the condition or illness of the patient, the technology, the value proposition, the adopters, the organization, the wider system, and the embedding and adaptation of the technology over time.

Stakeholders in different parts of a health care system may have differing inherent perspectives on key system functions in this complexity, depending on the role they have in the system. Glouberman and Mintzberg [9] observed that when managing changes, barriers need to be broken down between different perspectives on care (*curing*, *caring*, *costs*, and *community*) so that the resources of the overall system can be allocated more effectively. To navigate toward a responsible and sustainable future use of eHealth technology, it will be necessary to take on the challenge of understanding implementation complexity. This includes understanding the views of both the *back-end* stakeholders in an implementer role (the policy makers) and the *front-end* stakeholders in an end user role (the service users). Although several recent studies have identified factors that influence the perceptions of stakeholders in different populations [10-12], research on the experiences of stakeholders in a policy role is scarce, as is knowledge about the overlap between policy and public (service user) perspectives. The viewpoints of policy makers are seldom recorded in scientific publications despite their important role in the *outer context* of implementation [13-15].

Objectives

The aim of this study is to explore the perceptions concerning the barriers to and facilitators of the implementation of eHealth as perceived by policy makers and service users, and explore the ways in which their perceptions converge and differ. Ultimately, the study may contribute new knowledge to more effectively co-design the future implementation of digital health care technology.

Methods

Design

This study used a qualitative design in 2 steps, including interviews with 7 policy makers and 25 service users from the southeast of Sweden.

Participants and Interviews

The study includes primary data from 7 persons in a policy-making role and secondary data from 25 persons who had personal experience in an eHealth intervention.

The policy makers, employed in 4 different county councils in Southeast Sweden, were recruited by phone or email at the beginning of 2018. During a recruitment period of several months, a total of 23 persons representing different levels of policy makers in health care were contacted. Of these 23 persons, 14 (61%) declined to participate because of full agendas, and 2 (9%) did not answer the invitation.

The service user data were retrieved from 3 different eHealth intervention effectiveness studies [16-18]. Participants were provided written and oral information concerning the objectives and procedures of the study. The document informed the study participants that participation was voluntary and that they could stop at any moment with no reason given. All participants gave their informed consent.

Interviews with policy makers were conducted either face to face or by telephone, depending on their choice. The interview guide for policy makers was developed from the Swedish national eHealth policy document, with which participants were familiar [5]. The participants were asked to describe their perceptions and experiences with eHealth technology in their stakeholder role, their perceptions of the pros and cons of eHealth technology, their knowledge of policy documents related to eHealth technology, and their perceptions of factors influencing the implementation of eHealth in clinical practice.

Participants were interviewed individually either at their workplace or by telephone (according to their preference). Interviews lasted for a duration of 43 minutes to 63 minutes.

Service users were interviewed via telephone. In HeaDING (Heart Failure and Depression Using Internet Based–Cognitive Behavioural Therapy), service users were interviewed post intervention in 2015 [16]. All service users (n=27, with a median age of 69 years) who had been enrolled in internet-based cognitive behavioral therapy (iCBT) were invited to the interview study; of the 27 patients, 13 (48%) agreed to participate; of the 13 patients, 8 (51%) were men, and 5 (39%) were women. Interviews from 5 women and 5 men were used for the analysis in this study. Service users in DOHART (a research study concerning iCBT for persons with heart disease and depressive symptoms) [17] were interviewed post intervention between January and June 2017. A total of 35 service users with a mean age of 62 years were invited; of the 35 service users, 20 (57%) accepted to be interviewed; of the 20 participants, 11 (55%) were men, and 9 (45%) were women. In both HeaDING and DOHART studies, service users were included if they were receiving cardiovascular disease (CVD) treatment according to the current guidelines for heart failure, coronary artery disease, and atrial fibrillation and had at least mild depression (Patient Health Questionnaire ≥5 points). Interviews from 5 men and 5 women were used in this study. The service users in Heart And Insomnia Treatment Through Internet. (a research study concerning iCBT treatment for insomnia in patients with CVD) were interviewed before the intervention (late 2017) [18]. Service users had primary CVD as defined by diagnostic codes from the International Classification of Disease and sleep disorders. Diagnoses included myocardial infarction, angina pectoris, heart failure, atrial fibrillation or flutter, or arrhythmia. Service users were recruited from 6 primary care centers to participate in the randomized controlled trial. All participants in the intervention group (n=24, mean age 72 years) who had started their treatment were invited to participate in the interviews. As the study was ongoing at the time of this study, only 5 interviews (n=3, 60% men and n=2, 40% women) had been conducted, all of which were used for this study.

The primary focus of these interviews was on participants' experiences with the specific eHealth intervention in which they were enrolled; the focus on their general perceptions of factors influencing the implementation of eHealth in clinical care was secondary. Questions could concern their general perceptions of the applicability of the eHealth intervention to other users, the skills necessary for using the eHealth intervention, and the pros and cons of eHealth. The quality of the data varied from interview to interview. Where the service user interviews had relevant data for the purpose of this study, excerpts were analyzed. Some interviews were relevant in their entirety. Therefore, there was a wide range of interview duration (15-100 minutes).

Interviewers had clinical or research experience in cardiac nursing (service user interviews) or clinical and research experience in implementation research and health care (policy maker interviews), and all were experienced interviewers. All interviews were audio-recorded and transcribed verbatim.

Data Analysis

Implementation is the conduction, execution, or practice of a plan, a method, or any design, idea, model, specification, standard, or policy for doing something. Implementation science is "the scientific study of methods to promote the systematic uptake of research findings and other EBPs into routine practice, and, hence, to improve the quality and effectiveness of health services" [19]. The term *barrier* is used in implementation science to denote the challenges that may occur during the process of implementation and that may limit the effectiveness of the implementation strategy used. The term *facilitator* is used to denote a factor (condition) that may make implementing something easier. The analysis was performed in 2 steps. The first analysis (assorted analysis) explored the perceptions of each participant group separately, and the second analysis (supplementary analysis) explored aspects that needed to be understood further [20].

In the first step, the aim was to explore perceptions concerning barriers to and facilitators of the implementation of eHealth from the separate perspectives of policy makers on the one hand and service users on the other.

The data from policy makers and the data from service users were initially analyzed separately in each group by MN and JL in a process called *assorted analysis* [20] and adhered to the procedure of inductive qualitative content analysis as described by Hsieh and Shannon [21]. MN and JL read and reread the data several times using an inductive approach. They identified meaning units comprising words or sentences and, after subjecting these to a critical analysis, combined the content into separate sets of subcategories and categories. Barriers to and facilitators of the implementation of eHealth emerged in each analysis; these were the main categories within each group. A number of nonoverlapping and mutually excluding subcategories also emerged within the main categories for each group. To support trustworthiness, the results from these analyses were thereafter iteratively discussed in the wider author group until consensus was reached. The analyses were assisted by a computer software (NVivo 11 [QSR International]).

In the second step, a *supplementary analysis* was undertaken to explore aspects of the data that were not fully addressed in the first analysis [20]. This analysis (an analysis between groups), in which the perceptions of the 2 groups were compared and contrasted to explore differences and commonalities between policy makers and service users, also followed the procedure of inductive qualitative content analysis [21]. JL and MN undertook the initial analysis by iteratively examining the qualitative data in the subcategories representing the barriers and facilitators in each group of interviews, finding the points where they converged and differed. When necessary, the original data from the interviews with service users and policy makers were consulted to check for consistency and accuracy. To support trustworthiness, the results from these analyses were then iteratively discussed in the wider author group until consensus was reached.

Ethical Considerations

The study was performed in accordance with the Swedish ethical law and the Declaration of Helsinki. The regional ethics board approved the study (Dnr 2016/72-31; Dnr 2015/258-31, 2017/378-32; Dnr 2011.166/31).

Results

Overview

The participants in the policy maker group had a variety of roles ranging from elected politicians with a political responsibility for health care portfolios at the regional level (representing 3 different political parties) to administrators in a more operational capacity at regional or local levels. All had an explicit mission to implement information technology in health care. The participants (5/7, 71% women and 2/7, 29% men) were aged between 40 and 59 years and had work experience between 4 and 20 years in their present role. The participants in the service user group (9/25, 36% women and 16/25, 64% men) were aged 39 to 83 years and varied in educational background. Participants had various heart diseases: ischemic heart disease, heart failure, arrhythmia, or combinations thereof. All service users had an experience with eHealth, either having been introduced to, participating in, or having participated in an eHealth intervention.

Policy Makers

Overview

For policy makers, 5 barriers and 3 facilitators for the implementation of eHealth were found (Table 1).

Table 1. Main categories and subcategories with quotes from the qualitative content analysis (policy makers' perceptions of barriers to and facilitators of the implementation of eHealth interventions in health care).[a]

Subcategories	Quotes
Barriers	
Belief that not all service users will benefit from eHealth	"[eHealth technology]...will suit some people's needs, but different people need different things...I really think you should see it as complementary and not instead of [standard care]" [Policy maker 2]
Uncertainty about the work consequences for health care professionals	"We have staff working with the technical coordination...providing access and such...But about the human development...How you change the ways [health care] staff do their work...To be honest, I think we haven't been thinking about that a lot" [Policy maker 7]
Fear of problems with providing technical infrastructure	"We have to have a working internet in all parts of our region...If you can't get a sufficiently strong connection, that is going to be a problem" [Policy maker 2]
Perceived complexity of political decision-making	"There are at least 60 projects on-going [in our region] at the moment...We don't have an accepted way to make decisions: we can't compare different proposals, and we can't prioritize...we have no evaluation tool...We can't decide what is most important...is it what politicians think is best, or is it about how it affects our economy?...That's why it becomes a matter for everybody and also for nobody, because there isn't an obvious person or group who is responsible" [Policy maker 1]
Lack of extra resources for eHealth implementation	"I think about 'effect' from an economic perspective...all our practices are driven by economic considerations...It should be a smart solution...So if you use a technical solution, it should eradicate another cost" [Policy maker 3]
Facilitators	
Policy makers' conviction that eHealth is what citizens want	"For us in politics, the citizen perspective is the most important. To be able to access your own medical file, to sit at home and receive cognitive behavioral therapy. To be able to provide a chance to participate in the care process...it makes patients more participatory in their care" [Policy maker 4]
Policy makers' belief in eHealth solutions as beneficial for health care practice	"We expect to be more effective, and hope to use the time we gain to help more afflicted patients" [Policy maker 1]
Policy makers' belief in the importance of working toward eHealth implementation in their region	"Digitalization is part of our political mission. Our regional planning has been hampered because we are governed by a political minority which has made decision-making weak, so we have been a bit left behind. But my conviction is that this issue should be taken up to the highest political level...We have to show that we are in the game!" [Policy maker 3]

[a]Numbers in parentheses are participant codes.

Barriers Perceived by Policy Makers

Belief That Not All Service Users Benefit From eHealth

Although policy makers expressed very positive views about eHealth technology, some also discussed situations in which eHealth technology would be less beneficial: when professionals meet service users for the first time, when they prescribe psychiatric medication or assess the need for sick-leave, or when they treat the young or those service users with complex problems. Some service user characteristics (such as age and condition) were also described as barriers to eHealth technology; however, 1 policy maker thought that service users being uncomfortable with technology was the most important factor.

Uncertainty About the Work Consequences for Health Care Professionals

The process of changes in the direction of digitalization was described as uneven, as different work units were perceived to

be different in their way of tackling the issue and getting different support in the change process. A policy maker described that the changes that need to happen when introducing a new way of working in the clinic were not very well-developed and that implementing change was difficult.

Policy makers perceived that much effort was usually put into the technical implementation, whereas the development of the workforce was left to the work unit in question. Policy makers were uncertain about whether the planning for change and implementing eHealth technology in practice meant more work instead of less for work units. They perceived that staff could experience an uncertainty about the negative consequences of digitalization for their own employment and thought that staff needed better support for learning about technical solutions in the workplace.

Fear of Problems With Providing Technical Infrastructure

The incompatibility of the technical systems and differences in the *language* that systems use was perceived to lead to unnecessary difficulties in information sharing. Data sharing between health care systems was seen as a top priority by several participants, although some also raised the issue of data security.

Policy makers expressed that an important implementation barrier was a lack of knowledge about technical solutions and their real potential applicability in health care. They had many proposals about novel technical eHealth solutions flooding their workday and experienced difficulty managing the need to make implementation decisions in an informed and systematic way. There was uncertainty among policy makers about the organizational chain of responsibility for implementing the new technologies, the different actors concerned in the political system and the health care system, and how roles overlap and differ. As laws and regulations lag behind the current rapid technical development, there was a feeling of not being supported in decision-making (eg, when considering if and how to limit *internet doctors* from inappropriately diverting health care resources).

Lack of Extra Resources for eHealth Implementation

Despite the strong political impetus to implement eHealth solutions, policy makers perceived that no extra resources were available in the health care system for eHealth implementation. They perceived that implementation decisions at both the clinical and political levels were influenced by the lack of economical and personal resources, as they saw cost–benefit calculations as a key factor for implementation. They also found that decision-making in their political role was difficult; both the needs of the population and the solutions offered by technology companies were many and multifaceted. They found it hard to prioritize among solutions, as it was not always clear what the costs and gains would be in different cases. Policy makers expressed that they would prefer not to leave the development of digital tools exclusively to commercial agents with a profit agenda or to researchers with limited resources. They reasoned that extra resources should be allocated to implement technology in clinical everyday care and that public funds should engage more actively in supporting innovation, as resources are necessary to design and develop good digital products for health care use.

Facilitators Perceived by Policy Makers

The arguments that policy makers used to describe the benefits of eHealth solutions were many, and all policy makers extolled the advantages of digital care solutions for patients, staff, and society at large.

Policy Makers' Conviction That eHealth is What Citizens Want

Policy makers stressed that facilitating the accessibility to care for service users was an important driver for their work. Policy makers perceived that providing citizens with the opportunity to book their health care visits themselves, order their medication, and check their electronic health records on the web would guide health care to become more patient centered. Many perceived that citizens expect society to provide them with access to internet for professional examinations, treatment, and health monitoring, with the benefits of less traveling.

Policy Makers' Belief in eHealth Solutions as Beneficial for Health Care Practice

Perceptions were that eHealth technology could lead to advantages for staff by providing decision support. eHealth could lead to administrative effectiveness and better workflow; as more patients and families (in theory) managed their own health administration, staff could cater to those most in need of personal service.

Policy Makers' Belief in the Importance of Working Toward eHealth Implementation in Their Region

Some policy makers expressed that they saw the necessity to take an active role in politics and fulfill their ambition to push their region to the frontline of health care digitalization and implementation of eHealth. They expressed a wish to be among *the best*. In particular, building infrastructure was seen by policy makers as a crucial part of their political role, as populations in rural regions need broadband and other facilities to access eHealth programs. To reach that goal, policy makers saw a need to work together at the regional and local levels, involve service users, and collaborate with commercial companies and universities.

Service Users

Overview

For service users, 5 barriers to and 6 facilitators of the implementation of eHealth were found (Table 2).

Table 2. Main categories and subcategories with quotes from the qualitative content analysis (service users´ perceptions of barriers to and facilitators of the implementation of eHealth in health care).[a]

Subcategories	Quotes
Barriers	
Limitations in the capability of the service user	"You know, for me it was all this digitalisation. I had been better off if I could write on a paper and send it in. I thought it was rather complicated" [Service user 3]
eHealth is not always what the individual service user wants	"Yeah, when it comes to healthcare I think that you prefer a personal contact. Or so I think" [Service user 4]
eHealth is perceived as time consuming for the service user	"This a darn lot of time, for the patient, I mean it would take a lot less time if one had an appointment with someone and was there talking for an hour compared to reading a lot and fill out forms and find out answers and write answers" [Service user 3]
Mismatch of technology with service user needs	"Overall...Yes, I had, I must say had expected something else" [Service user 25]
Perceived lack of data protection	"I'm not happy about that at all...it's just a matter of how good a hacker you are..." [Service user 5]
Facilitators	
User-friendly design	"And even if you forget some things as time passed, you could always go back and refresh your memory. So that's an advantage" [Service user 22]
Matches skill set of service user	"I had an advantage in that I'm quite structured and used to work with structured materials on the computer" [Service user 22]
Provides a sense of privacy	"First you had a pass-word and then a single use code, that made me feel rather secure because they [designers of the web-platform] had thought about it" [Service user 2]
Personal feedback and staff support	"I had rather long questions and I got good answers and good feedback, I really appreciated that and without that it wouldn't work. It was a necessary and good complement [to the iCBT[b] program]" [Service user 1]
Flexible use of time	"If the alternative had been personal meetings, so that you had appointments a number of times, then naturally this [iCBT-program] has clear advantage timewise, because I have rather fully-booked work day, and I don't have to set more time aside for it" [Service user 26]
A credible sender	"I think this type of program, this type of treatment or healthcare, it's totally ok for me because you have a human that you can call and talk to, otherwise it wouldn't work" [Service user 3]

[a]Numbers in parentheses are participant codes.

[b]iCBT: internet-based cognitive behavioral therapy.

Barriers Perceived by Service Users

Limitations in the Capability of the Service User

Service users perceived that their health status could limit their ability to use eHealth, for example, because of memory problems. Some service users thought that other service users might hesitate to use digital technology in general and not only described technical problems with the digital solution itself but also the unfamiliarity with computer-based activity, for example, developing a smooth workflow on the web-based platform. They also found that the log-in process was complicated and would have preferred more automated feedback from the platform.

eHealth Is Not Always What the Individual Service User Wants

A feeling of being *tied to their computer* was perceived when working with the iCBT program. Service users could perceive that communication in writing with a health care professional was unfamiliar and sometimes uncomfortable for them, and those service users who were not used to reading long texts on screen expressed that they preferred to read from printouts.

Service users described the lack of face-to-face or direct audiovisual communication with the therapist as a barrier, as they perceived that communication in writing sometimes meant missing nuances in the communication. Some service users preferred traditional health care with face-to-face appointments and participation in support groups and therefore perceived eHealth as the second-best alternative.

eHealth Is Perceived as Time Consuming for the Service User

Although appreciative of the opportunity to use the eHealth technology when it suited them, service users would have liked to continue being able to contact health care by telephone. As this was perceived as increasingly difficult, they felt obligated to use digital solutions, which they described as complicated and time consuming. Some service users also felt that they had to take more responsibility for their treatment compared with traditional face-to-face treatment, for example, with weekly appointments. This was described as a barrier, especially for patients who experienced that their family or work situation required much time and attention.

Mismatch of Technology With Service User Needs

Service users described that they thought that the eHealth interventions in which they had participated did not match their health needs, that the intervention had not been provided at the right time, or that the content of the intervention did not meet their expectations or because it was not sufficiently tailored to their personal needs.

Perceived Lack of Data Protection

A barrier could also be uncertainty about personal integrity and information security. Service users who experienced uncertainty about what information could be extracted from different types of devices and telemonitoring expressed rather strong feelings, such as fear.

Facilitators Perceived by Service Users

User-Friendly Design

Service users perceived that a user-friendly design was important for their use of the eHealth intervention. An appropriate level of complexity in the program texts and communication processes (such as the option of ticking simple yes or no questions) was a feature. Although some service users described it as hard to write about health problems, some also described that the program relied on written communication and not real-time communication, as this meant that they had more time to consider how to communicate with health care staff in addressing sensitive topics. Another feature service users appreciated was the easy availability of the previous parts of the program.

Matches Skill Set of Service User

The service users in this study described themselves as relatively regular users of computers and digital services and perceived these experiences as facilitators of the use of an eHealth intervention. Specifically, previous experiences of, for example, working with texts, was perceived as a success factor in the intervention in which they were enrolled, as the iCBT program demanded a capacity for advanced reading and writing. For managing different types of slightly more complex processes, some service users needed help from family members.

Provides a Sense of Privacy

The service users in the study felt safe for the most part in the space that was provided for their eHealth activity and valued the direct contact it afforded with caregivers. When weighing the risk that an unauthorized person would get access to personal health-related information against the help they get via eHealth, they considered it worthwhile. This approach to privacy risks allowed participants to use eHealth; however, they also mentioned that the 2-factor authentication enabled their sense of privacy.

Personal Feedback and Staff Support

Service users perceived that personal contact with health care professionals before and during the eHealth intervention was an important facilitator. The study participants described a need for much information before the actual intervention began; they wanted information about the content of the eHealth program and also how to manage the computer-related tasks involved. Individual contact with health care personnel during ongoing interventions was perceived as important and even necessary. As eHealth was perceived as mostly managing health problems on one's own, being provided with an option to call someone for support was very valuable.

Flexible Use of Time

Many study participants perceived that their eHealth intervention took less time compared with, for example, physical visits within the care system. Among other things, service users appreciated that they did not have to set aside time for travel or look for parking and that participating in an eHealth intervention was easier than attending a traditional treatment with physical meetings as it was easier to fit into their daily schedule.

Service users perceived a greater degree of self-determination and awareness about when and how they conducted the treatment and appreciated feelings of being in control. For example, they could do exercises whenever they wanted and could read at their own pace.

A Credible Sender

Service users expressed that they perceived the sender of an eHealth intervention as reliable and credible, which was an implementation facilitator from their perspective. Service users described that they needed to feel that they could trust the organization *behind* the intervention. Public actors such as universities and public health care organizations were given greater credibility than private actors. Service users also described that it was important that they knew or experienced trust for the person responsible for their treatment and needed to know that they could get in touch with that person if needed.

Differences and Commonalities

Overview

In this section, we present the results of our supplementary analysis (Multimedia Appendix 1). In Multimedia Appendix 1, the boxes indicate subcategories from step 1, sorted by stakeholder group (policy makers and service users) and main category (barriers and facilitators). Arrows indicate similarities between stakeholder group perceptions, as expressed in the subcategories. The differences between the stakeholder subcategories are not marked.

Barriers

Policy makers perceived that a barrier to the implementation of eHealth was that not all service users benefit from eHealth. This was mirrored in the perceptions of service users, who highlighted the individual ability and preferences of service users as potential barriers.

Although policy makers perceived that there were implementation barriers in their work role (the complexity of decision-making, a lack of resources for health care digitalization, and fear of technical infrastructure problems), service users expressed that the design of the technology needed to be matched to their needs and that a perceived lack of data protection could be an implementation barrier to service users.

Both groups reflected on a possible redistribution of *work burden* between professionals and service users. Policy makers

expressed some uncertainty about the work consequences for health care professionals, and some service users perceived that they found themselves doing more of the professionals' work.

Facilitators

Policy makers and service users all perceived that the large impetus toward the implementation of eHealth should be considered an important facilitator. Policy makers were convinced about the public *pull* toward eHealth and believed that the importance of their policy mission would facilitate implementation.

Service users were more specific in describing which actual eHealth features of the technology would facilitate the implementation; not only should the eHealth design be user friendly and provide flexible use of time but also that it should provide a sense of privacy and a credible sender behind the technology.

Although policy makers believed that health care practice would benefit from eHealth by freeing up staff hours and saw that as an implementation driver, service users saw personal feedback and support from health care staff as an implementation facilitator.

Discussion

Principal Findings

This study shows that perceptions of policy makers and service users tend to converge concerning the perception of general societal impetus toward the adoption of eHealth and the promise of effectiveness through the flexibility of time use. The 2 groups also agreed that the implementation of eHealth may be hindered by the difficulties in tailoring eHealth to user differentiation and expressed that an implementation barrier could be the uncertainty about the ultimate distribution of care burden. The perceptions of policy makers and service users did not converge in other aspects. At the back end of implementation, policy makers perceived decision-making as complex, as technical and regulatory systems are lacking in relevance and detail, and knowledge and resources for implementation (ie, for building the eHealth infrastructure) are scarce. At the front end, service users highlighted the importance of the protection of digital data and trust in the sender.

In the following sections, we will use the domains in the *nonadoption, abandonment, spread, scale up, and sustainability* framework (framework for considering influences on the adoption, nonadoption, abandonment, spread, scale up, and sustainability of patient-facing health and care technologies) [8] to frame our discussion of the commonalities and differences in the perceived barriers to and facilitators of the implementation of eHealth between policy makers and service users.

In our study, both groups appeared to agree that the *condition and illness* of the service user is a factor to consider. They identified a common barrier in that eHealth is not a suitable or sufficient way of providing health care to every member of the general population. Although policy makers mainly focused on criteria such as high age and critical health as barriers, service users also identified the timing of eHealth interventions in their disease progress and their individual expectations of efficacy. The implementation of digital technology will always be more complex when the condition is poorly characterized, poorly understood, unpredictable, or high risk [8], and studies have shown that the medical and sociocultural criteria for patient selection for eHealth interventions are still in its infancy in many instances [22,23]. Greenhalgh et al [8] also reported that only a fraction of potential end users was assessed by their clinicians as *suitable* for an eHealth intervention. In the SARS-CoV-2 pandemic era, the use of digital tools has accelerated and spread more widely to many new areas of health care [24-26]. The question of whether current expectations about the general applicability of eHealth are, in fact, realistic remains to be seen. Recent publications note that despite societal pressures because of the pandemic, there is no extensive clinical experience of eHealth in cardiac care and highlight both the limitations and usefulness of digital solutions for managing CVD [27,28].

The *technology* itself has a prominent role in the perceptions of both groups. Service users expressed that the feeling of being in a digitally protected environment and trusting the *sender* facilitated implementation. Other research has confirmed that the authority of the author has been found to influence trust and credibility [29,30]. Although the policy makers in the study perceived that most of their citizens were interested in digital health innovation, they did not elaborate on the importance of cybersecurity and expressed that they had limited knowledge about the key features of different types of eHealth. Wozney et al [31] interviewed key informants who were influential in the adoption of technology for e-Mental health and found similar knowledge gaps. Swedish national laws and European laws are very clear about the responsibilities of the government and policy makers to protect citizens from harm through health data breaches [32,33]. In view of the need to accelerate the implementation of eHealth, Swedish authorities have recently devised an action plan. The main action targets in the plan are more focused attention to issues of data protection, the involvement of individual service users and health care staff, provision of information and knowledge, and promotion of collaboration around digital transformation [34]. A recent review focusing on the implementation of videoconferences in diverse areas of health care similarly found several challenges related to service development [35].

Notwithstanding the limited scientific evidence of the effectiveness of eHealth, policy makers and service users alike perceived that eHealth can enable more effective use of time in the management of health-related problems as a *value proposition*. However, one of the findings of the study was that both groups expressed uncertainty about the distribution of care burden. Service users saw eHealth as a way of making health care fit in their daily schedule and meeting their needs. Other studies confirm these perceptions of benefits for service users [12,36-38]. Our study showed that service users could also perceive their eHealth intervention as time consuming; however, they expressed concern about the shifting of roles and responsibilities (and workload) from health care practitioners to service users and their caregivers. This shift in the burden of care has been reported to change the work required of patients and be a potential barrier for the use of eHealth, which may lead

to nonuse by service users [8,39]. Policy makers primarily saw eHealth as potentially beneficial for health care practice, for example, a way of reducing care tasks among health care professionals. In a recent systematic review of primary health care workers' perceptions and experiences of mobile health (mobile technology and communication tools such as smartphones, tablets, PDAs, and wearable devices such as smartwatches for the prevention, promotion, treatment, and maintenance of health, information, and data collection), results were shown to be rather more mixed. Although health workers recognized some time benefits, they saw mobile health as slow and time consuming in some cases and as creating more work [40]. The findings of Greenhalgh et al [8] point to the real risk of some staff simply not engaging with the program or using the technology. In a recent review by Drissi et al [41], only a small part of the included studies provided an empirical evaluation of the reported digital interventions to assist health care workers in mental care, and half of the studies listed challenges and limitations related to the adoption of the reported interventions.

Policy makers and service users in the study both agreed that eHealth potentially allows patients to be more empowered to participate in their own care. Similarly, a study by Whittaker [42] showed that policy makers, administrators, and organizational leaders in the United States viewed eHealth technology as potentially transformative for health care. Other research shows that patient-facing technologies may not only inform, educate, and empower but also sometimes be misinterpreted and cause distress [8,43]. In this study, service users expressed concerns that implementing eHealth would mean less human contact with health care professionals, which is something they saw as an important quality marker for health care. A recent systematic review found that there were few studies that identified ethical issues associated with telehealth practice [44]. A systematic review by Parker et al [45] found that differences in socioeconomic factors and gender were associated with the use of remote consultations and internet-based consultations in general practice. In their systematic review of the use of eHealth technology in cardiac care, Harky et al [46] highlighted the need for further studies exploring how staff–service user communications answer the call for service user empowerment.

The 2 stakeholder groups differed in their perceptions of factors related to *adopters*. One of the important results from our analysis was that policy makers did not explicitly show an awareness of the features of eHealth that service users perceived as important facilitators of their engagement with the technology. Although policy makers focused on the back-end challenges with eHealth, for example, the difficulty of providing regional broadband coverage, service users were mostly concerned about the interface and their own ability to work, read, and use the digital solution. Research has shown that the information quality and the demands of the technology on service users' computer literacy skills and other individual competencies are very important [47,48]. However, tailored solutions that cover variable competencies come at a considerable monetary cost [8]. As the policy makers in the study expressed the real-life need to align their implementation decisions to a health care system with a nonexpanding budget, individual tailoring of technologies to differing needs will realistically be limited.

In an *organizational and wider system* context, the policy makers in the study had an ambition toward making their region's digitalization compare favorably with other parts of the country. This raises the question of which priorities will be set in the near future: regional interests or collaboration across regional borders. Policy makers perceived uncertainty in their decision-making, as there was a lack of regulatory support to guide their decision-making and no extra resources for implementing the technologies. The Swedish vision of national digitalization of health care by the year 2025 [5] stands in stark contrast to policy makers' perceptions of decisional difficulties, pointing to the need for more coherent national planning, strategic reimbursement, and regulatory support. Other research also reports key barriers in micro- and mesolevel contexts of health care organizations; however, more research is needed that focuses explicitly on policy development and implementation planning [29,49,50].

Neither policy makers nor service users reflected on the *embedding of technology in health care systems* and the *adaptation of technology over time.* Although the research literature on the use of eHealth has exploded during the SARS-CoV-2 pandemic, not many studies have used more rigorous theoretical tools to study the long-term issues of using eHealth technologies.

The World Health Organization highlights that there is a need to go beyond the small-scale implementation of pilot projects focusing on the evidence of feasibility and effect to a more extensive exploration of the infrastructure needed to scale up and sustain digital health technology as a global issue in the future. *Scaling up* comprises deliberate efforts to increase the impact of innovations that were successfully tested in pilot or experimental projects so as to benefit more people. Policy and program development on a lasting basis may be fostered through government adoption, commercial adoption, or hybrid models [51]. Stable and secure financial and technical resources and enduring partnerships are the foundation for sustainability, in addition to the capacity to continually adapt the product to meet the demands of service users and the ever-changing operational environment [52]. However, an understanding building on new insights may take some time. Greenhalgh et al [8] proposed that some of the barriers to achieving better eHealth implementation may involve multiple strategies at the organizational and societal level, including a revision of laws and a gradual "shared sense-making through reflexive monitoring and analysis of critical events or issues"

In any event, outcomes from eHealth solutions need reasonably be superior or at least comparable with traditional health care practices in terms of safety, timeliness, equity, effectiveness, efficiency, and patient centeredness. The SARS-CoV-2 pandemic has made practitioners and researchers more aware that although eHealth solutions have a great attraction, there are many aspects that remain to be explored. New research findings will likely fuel the discussion on the digital transformation of society and health care. Understanding the

complexities involved in this transformation is a challenge not only for the public but also for policy makers at all levels; setting an even more proactive agenda for the solid and high-quality implementation of eHealth will be necessary. To make well-based policy decisions concerning the implementation of eHealth, policy makers will need to be informed by research that takes the complexity of implementation into account and pays more systematic attention to implementation outcomes that ensure the quality of care.

Before attempting to implement various eHealth solutions, implementation studies are not only needed to identify and address numerous barriers to implementation and understand how interventions can be sustainably translated from research into clinical practice but also to evaluate outcomes with standardized evaluation methods and instruments that are lacking at present [53].

Strengths and Limitations of the Study

As the study was conducted in prepandemic conditions, the incentive to use eHealth was not as strong as it has since become, and perceptions about using eHealth technology will likely have evolved in both stakeholder groups.

The participants in both the policy maker and the service user group had a variety of demographic characteristics, which indicates some representativeness. The perspectives of policy makers, who are an underrepresented group in health care research, contribute to a broader insight into barriers to and facilitators of the implementation of eHealth. The scope of the service users' perceptions of eHealth implementation, in general, was possibly somewhat limited by the fact that the focus in the interviews was on their experiences with the specific eHealth intervention (e-Mental health in cardiac care) in which they were enrolled. However, as the service users actively participated in eHealth interventions, they were more knowledgeable about eHealth than other members of the public, which benefited the study.

The authors' group expertise, with competencies in implementation science, extensive clinical experience in health care, and experience of qualitative methodology, contributes to methodological trustworthiness.

Conclusions and Clinical Implications

This study provides previously unavailable information about key informant perspectives on eHealth implementation. The study not only shows that both policy makers and service users perceive an impetus toward the implementation of eHealth but also that there are differences in views concerning implementation challenges and that policy makers do not perceive the barriers and facilitators in the same way as service users do. Dissonant perceptions about a *new distribution of workload* emerged; although policy makers see eHealth as potentially freeing up staff hours, service users highlight their need for data security, feedback, and support and may sometimes even see eHealth as the lesser alternative.

To be able to gear policy making to match service user needs and preferences, future research should explore avenues toward a more effective knowledge exchange regarding eHealth implementation between different stakeholder groups and systematic use of implementation science frameworks. Further research is needed to clarify the perspectives of health care staff and their clinical leaders on the sustainable implementation of eHealth in clinical practice and how their perceptions overlap with the views expressed by the stakeholders in this paper.

Acknowledgments

The authors would like to thank Professor Per Nilsen for constructive discussions.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Visual illustration of the qualitative content analyses.
[PNG File , 185 KB - jmir_v24i1e28870_app1.png]

References

1. State of Health in the EU: shift to prevention and primary care is the most important trend across countries. European Commission. Luxembourg: Publications Office of the European Union; 2019. URL: https://ec.europa.eu/commission/presscorner/detail/en/IP_19_6336 [accessed 2020-06-11]
2. World Health Organization. Global Diffusion of eHealth: Making Universal Health Coverage Achievable: Report of the Third Global Survey on eHealth. Geneva: World Health Organization; 2016. URL: https://apps.who.int/iris/handle/10665/252529 [accessed 2021-12-30]
3. Black AD, Car J, Pagliari C, Anandan C, Cresswell K, Bokun T, et al. The impact of eHealth on the quality and safety of health care: a systematic overview. PLoS Med 2011 Jan;8(1):e1000387 [FREE Full text] [doi: 10.1371/journal.pmed.1000387] [Medline: 21267058]
4. Flodgren G, Rachas A, Farmer AJ, Inzitari M, Shepperd S. Interactive telemedicine: effects on professional practice and health care outcomes. Cochrane Database Syst Rev 2015;9:CD002098. [doi: 10.1002/14651858.CD002098.pub2] [Medline: 26343551]

5. Swedish Government and Swedish Association of Local Authorities and Regions (SALAR). Vision for eHealth 2025 : follow-up 2019. E-Hälsa 2025.: E-hälsomyndigheten; 2020. URL: https://ehalsa2025.se/wp-content/uploads/2021/02/Follow-up-2019_Vision-e-health-2025.pdf [accessed 2020-06-11]
6. Mair FS, May C, O'Donnell C, Finch T, Sullivan F, Murray E. Factors that promote or inhibit the implementation of e-health systems: an explanatory systematic review. Bull World Health Organ 2012 May 1;90(5):357-364 [FREE Full text] [doi: 10.2471/BLT.11.099424] [Medline: 22589569]
7. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: a systematic review of systematic reviews (an update). Implement Sci 2016 Oct 26;11(1):146 [FREE Full text] [doi: 10.1186/s13012-016-0510-7] [Medline: 27782832]
8. Greenhalgh T, Wherton J, Papoutsi C, Lynch J, Hughes G, A'Court C, et al. Beyond adoption: a new framework for theorizing and evaluating nonadoption, abandonment, and challenges to the scale-up, spread, and sustainability of health and care technologies. J Med Internet Res 2017 Nov 01;19(11):e367 [FREE Full text] [doi: 10.2196/jmir.8775] [Medline: 29092808]
9. Glouberman S, Mintzberg H. Managing the care of health and the cure of disease - Part I: Differentiation. Ovid. 2001. URL: https://ovidsp.dc2.ovid.com/ovid-b/ [accessed 2021-12-09]
10. O'Connor S, Hanlon P, O'Donnell CA, Garcia S, Glanville J, Mair FS. Understanding factors affecting patient and public engagement and recruitment to digital health interventions: a systematic review of qualitative studies. BMC Med Inform Decis Mak 2016 Sep 15;16(1):120 [FREE Full text] [doi: 10.1186/s12911-016-0359-3] [Medline: 27630020]
11. Swinkels IC, Huygens MW, Schoenmakers TM, Nijeweme-D'Hollosy WO, van Velsen L, Vermeulen J, et al. Lessons learned from a living lab on the broad adoption of eHealth in primary health care. J Med Internet Res 2018 Mar 29;20(3):e83 [FREE Full text] [doi: 10.2196/jmir.9110] [Medline: 29599108]
12. Slater H, Campbell JM, Stinson JN, Burley MM, Briggs AM. End user and implementer experiences of mhealth technologies for noncommunicable chronic disease management in young adults: systematic review. J Med Internet Res 2017 Dec 12;19(12):e406 [FREE Full text] [doi: 10.2196/jmir.8888] [Medline: 29233804]
13. Damschroder LJ, Aron DC, Keith RE, Kirsh SR, Alexander JA, Lowery JC. Fostering implementation of health services research findings into practice: a consolidated framework for advancing implementation science. Implement Sci 2009;4:50 [FREE Full text] [doi: 10.1186/1748-5908-4-50] [Medline: 19664226]
14. Vassilev I, Rowsell A, Pope C, Kennedy A, O'Cathain A, Salisbury C, et al. Assessing the implementability of telehealth interventions for self-management support: a realist review. Implement Sci 2015;10:59 [FREE Full text] [doi: 10.1186/s13012-015-0238-9] [Medline: 25906822]
15. Varsi C, Nes LS, Kristjansdottir OB, Kelders SM, Stenberg U, Zangi HA, et al. Implementation strategies to enhance the implementation of eHealth programs for patients with chronic illnesses: realist systematic review. J Med Internet Res 2019 Sep 27;21(9):e14255 [FREE Full text] [doi: 10.2196/14255] [Medline: 31573934]
16. Lundgren JG, Dahlström O, Andersson G, Jaarsma T, Kärner KA, Johansson P. The effect of guided web-based cognitive behavioral therapy on patients with depressive symptoms and heart failure: a pilot randomized controlled trial. J Med Internet Res 2016 Aug 03;18(8):e194 [FREE Full text] [doi: 10.2196/jmir.5556] [Medline: 27489077]
17. Johansson P, Westas M, Andersson G, Alehagen U, Broström A, Jaarsma T, et al. An internet-based cognitive behavioral therapy program adapted to patients with cardiovascular disease and depression: randomized controlled trial. JMIR Ment Health 2019 Oct 03;6(10):e14648 [FREE Full text] [doi: 10.2196/14648] [Medline: 31584000]
18. Siebmanns S, Johansson P, Ulander M, Johansson L, Andersson G, Broström A. The effect of nurse-led internet-based cognitive behavioural therapy for insomnia on patients with cardiovascular disease: a randomized controlled trial with 6-month follow-up. Nurs Open 2021 Feb 20;8(4):1755-1768 [FREE Full text] [doi: 10.1002/nop2.817] [Medline: 33609425]
19. Eccles MP, Mittman BS. Welcome to Implementation Science. Implementation Sci 2006 Feb 22;1(1):1. [doi: 10.1186/1748-5908-1-1]
20. Heaton J. Secondary analysis of qualitative data: an overview. Histor Soc Res / Historische Sozialforschung 2008;33(3):33-45 [FREE Full text] [doi: 10.12759/hsr.33.2008.3.33-45]
21. Hsieh H, Shannon SE. Three approaches to qualitative content analysis. Qual Health Res 2005 Nov;15(9):1277-1288. [doi: 10.1177/1049732305276687] [Medline: 16204405]
22. Joseph V, West RM, Shickle D, Keen J, Clamp S. Key challenges in the development and implementation of telehealth projects. J Telemed Telecare 2011;17(2):71-77. [doi: 10.1258/jtt.2010.100315] [Medline: 21097563]
23. Paré G, Moqadem K, Pineau G, St-Hilaire C. Clinical effects of home telemonitoring in the context of diabetes, asthma, heart failure and hypertension: a systematic review. J Med Internet Res 2010;12(2):e21 [FREE Full text] [doi: 10.2196/jmir.1357] [Medline: 20554500]
24. Hincapié MA, Gallego JC, Gempeler A, Piñeros JA, Nasner D, Escobar MF. Implementation and usefulness of telemedicine during the COVID-19 pandemic: a scoping review. J Prim Care Community Health 2020;11:2150132720980612 [FREE Full text] [doi: 10.1177/2150132720980612] [Medline: 33300414]
25. Lieneck C, Weaver E, Maryon T. Outpatient telehealth implementation in the united states during the COVID-19 global pandemic: a systematic review. Medicina (Kaunas) 2021 May 09;57(5):462 [FREE Full text] [doi: 10.3390/medicina57050462] [Medline: 34065050]

26. Gunasekeran DV, Tseng RM, Tham Y, Wong TY. Applications of digital health for public health responses to COVID-19: a systematic scoping review of artificial intelligence, telehealth and related technologies. NPJ Digit Med 2021 Feb 26;4(1):40 [FREE Full text] [doi: 10.1038/s41746-021-00412-9] [Medline: 33637833]
27. Kaushik KA, Patel S, Dubey K. Digital cardiovascular care in COVID-19 pandemic: a potential alternative? J Card Surg 2020 Dec;35(12):3545-3550 [FREE Full text] [doi: 10.1111/jocs.15094] [Medline: 33040399]
28. Ajibade A, Younas H, Pullan M, Harky A. Telemedicine in cardiovascular surgery during COVID-19 pandemic: a systematic review and our experience. J Card Surg 2020 Oct;35(10):2773-2784 [FREE Full text] [doi: 10.1111/jocs.14933] [Medline: 32881081]
29. Sbaffi L, Rowley J. Trust and credibility in web-based health information: a review and agenda for future research. J Med Internet Res 2017 Jun 19;19(6):e218 [FREE Full text] [doi: 10.2196/jmir.7579] [Medline: 28630033]
30. Westas M, Lundgren J, Mourad G, Neher M, Johansson P. How healthcare professionals in cardiac care address depressive symptoms: experiences of patients with cardiovascular disease. J Cardiovasc Nurs 2020 Feb 21;36(4):340-348. [doi: 10.1097/JCN.0000000000000669] [Medline: 32084077]
31. Wozney L, Newton AS, Gehring ND, Bennett K, Huguet A, Hartling L, et al. Implementation of eMental Health care: viewpoints from key informants from organizations and agencies with eHealth mandates. BMC Med Inform Decis Mak 2017 Jun 02;17(1):78 [FREE Full text] [doi: 10.1186/s12911-017-0474-9] [Medline: 28577543]
32. Regulation (EU) 2016/679 of the European Parliament and of the Council. European Union - General Data Protection Regulation. 2016. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj [accessed 2020-10-01]
33. Patient Safety Act (2010: 659). Swedish Patient Safety Law. 2020. URL: https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/patientsakerhetslag-2010659_sfs-2010-659 [accessed 2020-11-01]
34. Swedish Government and Swedish Association of Local Authorities and Regions (SALAR). Genomförandeplan 2020–2022 Bilaga till Strategidokument Vision e-hälsa 2025. E-Hälsa 2025.: E-hälsomyndigheten; 2021 Jun. URL: https://ehalsa2025.se/visionen/genomforandeplan/ [accessed 2021-09-01]
35. James HM, Papoutsi C, Wherton J, Greenhalgh T, Shaw SE. Spread, scale-up, and sustainability of video consulting in health care: systematic review and synthesis guided by the NASSS framework. J Med Internet Res 2021 Jan 26;23(1):e23775 [FREE Full text] [doi: 10.2196/23775] [Medline: 33434141]
36. Struthers A, Charette C, Bapuji SB, Winters S, Ye X, Metge C, et al. The acceptability of e-mental health services for children, adolescents, and young adults: a systematic search and review. Can J Commun Ment Health 2015 Jul 01;34(2):1-21. [doi: 10.7870/cjcmh-2015-006]
37. Thiyagarajan A, Grant C, Griffiths F, Atherton H. Exploring patients' and clinicians' experiences of video consultations in primary care: a systematic scoping review. BJGP Open 2020;4(1):bjgpopen20X101020 [FREE Full text] [doi: 10.3399/bjgpopen20X101020] [Medline: 32184212]
38. Tully L, Case L, Arthurs N, Sorensen J, Marcin JP, O'Malley G. Barriers and facilitators for implementing paediatric telemedicine: rapid review of user perspectives. Front Pediatr 2021;9:630365. [doi: 10.3389/fped.2021.630365] [Medline: 33816401]
39. Gilbert A, Jones J, Jaggi A, May C. Use of virtual consultations in an orthopaedic rehabilitation setting: how do changes in the work of being a patient influence patient preferences? A systematic review and qualitative synthesis. BMJ Open 2020 Sep 16;10(9):e036197 [FREE Full text] [doi: 10.1136/bmjopen-2019-036197] [Medline: 32938591]
40. Odendaal WA, Watkins JA, Leon N, Goudge J, Griffiths F, Tomlinson M, et al. Health workers' perceptions and experiences of using mHealth technologies to deliver primary healthcare services: a qualitative evidence synthesis. Cochrane Database Syst Rev 2020 Mar 26;3:CD011942 [FREE Full text] [doi: 10.1002/14651858.CD011942.pub2] [Medline: 32216074]
41. Drissi N, Ouhbi S, Marques G, de la Torre Díez I, Ghogho M, Idrissi MA. A systematic literature review on e-mental health solutions to assist health care workers during COVID-19. Telemed J E Health 2020 Sep 22:594-602. [doi: 10.1089/tmj.2020.0287] [Medline: 32970532]
42. Whittaker R. Issues in mHealth: findings from key informant interviews. J Med Internet Res 2012;14(5):e129 [FREE Full text] [doi: 10.2196/jmir.1989] [Medline: 23032424]
43. de Vries AE, van der Wal MH, Nieuwenhuis MM, de Jong RM, van Dijk RB, Jaarsma T, et al. Health professionals' expectations versus experiences of internet-based telemonitoring: survey among heart failure clinics. J Med Internet Res 2013;15(1):e4 [FREE Full text] [doi: 10.2196/jmir.2161] [Medline: 23305645]
44. Keenan AJ, Tsourtos G, Tieman J. The value of applying ethical principles in telehealth practices: systematic review. J Med Internet Res 2021 Mar 30;23(3):e25698 [FREE Full text] [doi: 10.2196/25698] [Medline: 33783366]
45. Parker RF, Figures EL, Paddison CA, Matheson JI, Blane DN, Ford JA. Inequalities in general practice remote consultations: a systematic review. BJGP Open 2021 Jun;5(3):BJGPO.2021.0040 [FREE Full text] [doi: 10.3399/BJGPO.2021.0040] [Medline: 33712502]
46. Harky A, Adan A, Mohamed M, Elmi A, Theologou T. Technology and cardiovascular diseases in the era of COVID-19. J Card Surg 2020 Dec;35(12):3551-3554 [FREE Full text] [doi: 10.1111/jocs.15096] [Medline: 33040461]
47. Su JJ, Yu DS, Paguio JT. Effect of eHealth cardiac rehabilitation on health outcomes of coronary heart disease patients: a systematic review and meta-analysis. J Adv Nurs 2020 Mar;76(3):754-772. [doi: 10.1111/jan.14272] [Medline: 31769527]

48. Baumel A, Kane JM. Examining predictors of real-world user engagement with self-guided ehealth interventions: analysis of mobile apps and websites using a novel dataset. J Med Internet Res 2018 Dec 14;20(12):e11491 [FREE Full text] [doi: 10.2196/11491] [Medline: 30552077]
49. Batterham PJ, Sunderland M, Calear AL, Davey CG, Christensen H, Teesson M, et al. Developing a roadmap for the translation of e-mental health services for depression. Aust N Z J Psychiatry 2015 Apr 23:776-784. [doi: 10.1177/0004867415582054] [Medline: 25907269]
50. Meurk C, Leung J, Hall W, Head BW, Whiteford H. Establishing and governing e-mental health care in australia: a systematic review of challenges and a call for policy-focussed research. J Med Internet Res 2016 Jan 13;18(1):e10 [FREE Full text] [doi: 10.2196/jmir.4827] [Medline: 26764181]
51. World Health Organization. The MAPS Toolkit: mHealth Assessment and Planning for Scale. Geneva: World Health Organization; 2015:1-94.
52. Simmons R, Fajans P, Ghiron L. Introduction. In: Simmons R, Fajans P, Ghiron L, editors. Scaling Up Health Service Delivery: From Pilot Innovations to Policies and Programmes. Geneva: World Health Organization; 2007:vii-xvii.
53. Vis C, Bührmann L, Riper H, Ossebaard HC. Health technology assessment frameworks for eHealth: a systematic review. Int J Technol Assess Health Care 2020 Apr 16;36(3):204-216. [doi: 10.1017/s026646232000015x]

Abbreviations

CVD: cardiovascular disease
iCBT: internet-based cognitive behavioral therapy
HeaDING: Heart Failure and Depression Using Internet Based–Cognitive Behavioural Therapy

Edited by R Kukafka; submitted 18.03.21; peer-reviewed by M Newton, E Wang, B Cotner; comments to author 31.05.21; revised version received 01.10.21; accepted 14.10.21; published 28.01.22.

Please cite as:
Neher M, Nygårdh A, Broström A, Lundgren J, Johansson P
Perspectives of Policy Makers and Service Users Concerning the Implementation of eHealth in Sweden: Interview Study
J Med Internet Res 2022;24(1):e28870
URL: https://www.jmir.org/2022/1/e28870
doi:10.2196/28870
PMID:35089139

Original Paper

Sequential Data–Based Patient Similarity Framework for Patient Outcome Prediction: Algorithm Development

Ni Wang[1,2], BS; Muyu Wang[1,2], BS; Yang Zhou[3], BS; Honglei Liu[1,2], PhD; Lan Wei[4], PhD; Xiaolu Fei[4], PhD; Hui Chen[1,2], PhD

[1]School of Biomedical Engineering, Capital Medical University, Beijing, China

[2]Beijing Advanced Innovation Center for Big Data-based Precision Medicine, Capital Medical University, Beijing, China

[3]Department of Epidemiology and Biostatistics, Institute of Basic Medical Sciences Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing, China

[4]Information Center, Xuanwu Hospital, Capital Medical University, Beijing, China

Corresponding Author:
Hui Chen, PhD
School of Biomedical Engineering
Capital Medical University
No.10, Xitoutiao, You An Men, Fengtai District
Beijing, 100069
China
Phone: 86 010 8391 1545
Email: chenhui@ccmu.edu.cn

Abstract

Background: Sequential information in electronic medical records is valuable and helpful for patient outcome prediction but is rarely used for patient similarity measurement because of its unevenness, irregularity, and heterogeneity.

Objective: We aimed to develop a patient similarity framework for patient outcome prediction that makes use of sequential and cross-sectional information in electronic medical record systems.

Methods: Sequence similarity was calculated from timestamped event sequences using edit distance, and trend similarity was calculated from time series using dynamic time warping and Haar decomposition. We also extracted cross-sectional information, namely, demographic, laboratory test, and radiological report data, for additional similarity calculations. We validated the effectiveness of the framework by constructing k–nearest neighbors classifiers to predict mortality and readmission for acute myocardial infarction patients, using data from (1) a public data set and (2) a private data set, at 3 time points—at admission, on Day 7, and at discharge—to provide early warning patient outcomes. We also constructed state-of-the-art Euclidean-distance k–nearest neighbor, logistic regression, random forest, long short-term memory network, and recurrent neural network models, which were used for comparison.

Results: With all available information during a hospitalization episode, predictive models using the similarity model outperformed baseline models based on both public and private data sets. For mortality predictions, all models except for the logistic regression model showed improved performances over time. There were no such increasing trends in predictive performances for readmission predictions. The random forest and logistic regression models performed best for mortality and readmission predictions, respectively, when using information from the first week after admission.

Conclusions: For patient outcome predictions, the patient similarity framework facilitated sequential similarity calculations for uneven electronic medical record data and helped improve predictive performance.

(J Med Internet Res 2022;24(1):e30720) doi:10.2196/30720

KEYWORDS

patient similarity; electronic medical records; time series; acute myocardial infarction; natural language processing; machine learning; deep learning; outcome prediction; informatics; health data

Introduction

In recent years, personalized medicine and clinical decision-making support have become popular issues and hot research fields such as modeling with electronic medical records to assist clinicians in diagnosing diseases [1-3], predicting length of hospital stay [4,5], and predicting patient death and other outcomes [4,6-8]. Because electronic medical record data accumulate quickly, sufficient data exist for conducting data-driven studies, big data mining, and constructing predictive models. Using patient similarity measures calculated from electronic medical record data to select study cohorts for building personalized models has improved predictive performances [9,10].

Previous studies [11-14] have demonstrated the effectiveness of personalized predictive models. Wang et al [11,12] used similarity-based models to predict diabetes and liver disease risk. Li et al [13] successfully identified 3 distinct subgroups of type 2 diabetes based on the calculated patient similarity. Wang et al [14] derived a local spline regression-based method for patient embedding and patient similarity measurement to predict cardiovascular disease risk. However, these studies [11-14] merely evaluated patient similarity based on cross-sectional information, rather than using the complete longitudinal information stored in the electronic medical record system. For a hospitalized patient, the longitudinal information represents the clinical trajectory from admission to discharge; it may include a series of clinical events performed on a patient and multiple laboratory tests. Longitudinal data should be better than cross-sectional data in predicting patients' outcomes due to the rich information on medical behavior and disease progression. Thus, we can assume that longitudinal information in conjunction with patient similarity measurements will further improve outcome prediction, which will facilitate the move toward personalized medicine.

Unfortunately, as is typical of real-world data, electronic medical record data are usually heterogeneous, irregular, and uneven, which presents challenges for modeling and measuring similarity [15]. These problems are more severe for sequential information than they are for cross-sectional information. Thus, many researchers transform longitudinal data into static data. Lee et al [16] extracted various clinical and vital signs during the first 24-hour intensive care unit stay. These longitudinal variables were transformed into static data by calculating the minimum and maximum value for further patient similarity measurement based on the cosine similarity metric. Ng et al [17] used a feature vector representation method to aggregate longitudinal patient data by calculating counts for categorical variables (diagnoses, medications, and procedures) and arithmetic means for numeric lab test data. Sun et al [18] represented 2-hour temporal data for each patient by computing the means and variances or wavelet coefficients.

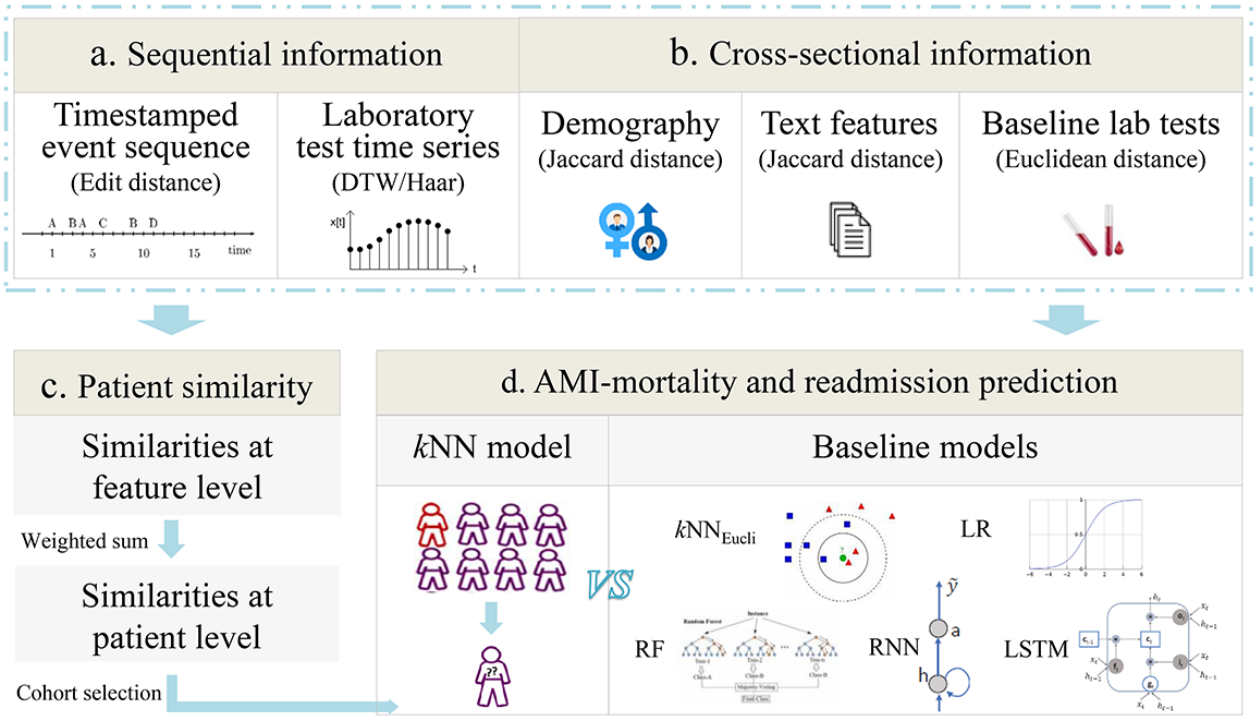
Because few analyses [15-18] have taken event sequences into consideration for similarity measurements, we aimed to develop a new framework for patient similarity measurement that can make use of cross-sectional information and 2 types of sequential information (series of clinical events and multiple laboratory tests) to predict patient outcomes.

Methods

Overview

In China, the number of patients with acute myocardial infarction is expected to increase from 8 million in 2010 to 23 million in 2030 [19], which usually has a high risk of all-cause in-hospital mortality or readmission, due to unexpected acute myocardial infarction, after discharge. The accurate prediction of these would allow better prognosis and timely intervention. Thus, we focused on the prediction of all-cause in-hospital mortality and unexpected acute myocardial infarction-readmission after discharge of patients with acute myocardial infarction at 3 time points during hospitalization (at admission, on Day 7, and at discharge). Each patient's clinical trajectory comprised a series of clinical processes (timestamped event sequence) and multiple laboratory tests (time series data) from electronic medical record data. We calculated similarities for both sequential and cross-sectional information and constructed similarity-based models for each time point (Figure 1).

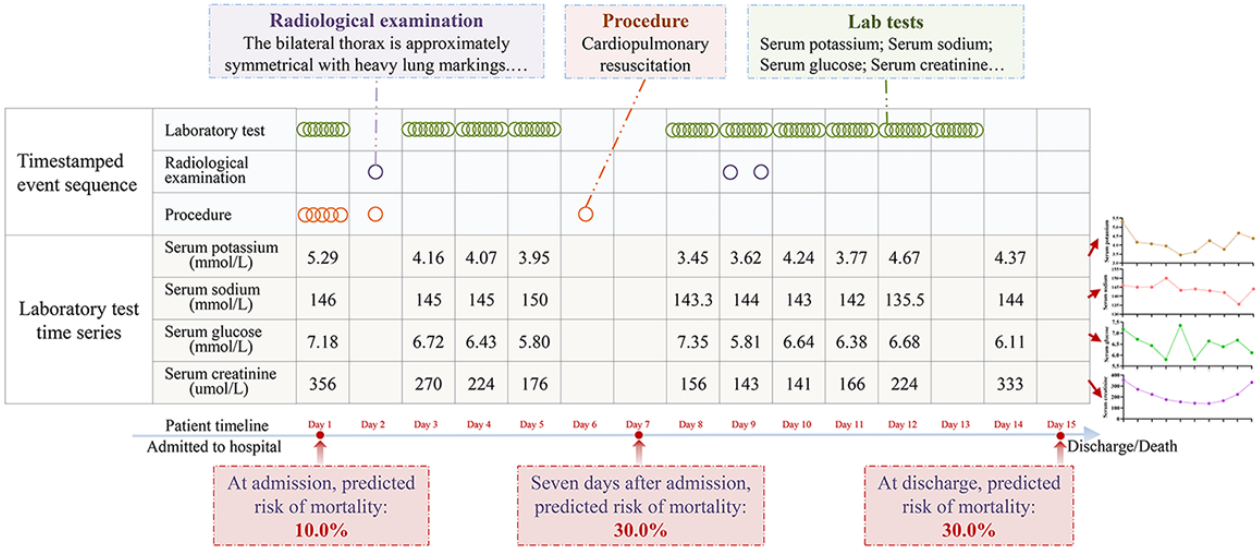
Figure 1. Study workflow: (a) Sequential similarity calculation for timestamped event sequence and time series data, (b) similarity calculation for cross-sectional information, (c) patient similarity measurement based on the weighted sum of similarities calculated in parts a and b, and (d) validation. AMI: acute myocardial infarction; kNN: k-nearest neighbors based on the proposed patient similarity measurement; kNN_{Eucli}: k-nearest neighbors based on the Euclidean distance; LR: logistic regression; RF: random forest; RNN: recurrent neutral network; LSTM: long short-term memory network; DTW: dynamic time warping.



Similarity for Sequential Information

Both timestamped event sequence and laboratory test time series data were used to calculate sequential similarity. Laboratory tests data contributed to sequence similarity calculations, and trend similarity calculations from multiple test values, simultaneously. Figure 2 shows an example of a patient's clinical trajectory.

Figure 2. A case study of a patient's clinical trajectory. All clinical events including laboratory tests, radiological examinations, and procedures are listed sequentially according to patient timeline. The multiple values of each laboratory test comprise the time series data shown in the line chart on the right side of the figure.



Similarity for Event Sequence

An event was a clinical process performed on a patient, such as a serum glucose test, a radiological examination (eg, color sonography), or a procedure (eg, percutaneous coronary intervention). For a series of clinical events, with timestamped information, an event sequence r for the patient comprised pairs (e_i, t_i), where e_i was the ith clinical event for a patient and t_i was

the time point (day) on which the event occurred. Within an event sequence, event e_i was placed before event e_j if t_i was earlier than t_j in the patient timeline. Two events were placed alphabetically if they were performed on the same day.

The edit distance was used to calculate the similarity between 2 event sequences based on how much work was needed to transform one sequence into the other [20,21]. Operations—insertion, deletion, and substitution—were used to change sequence r_1 into r_2. For the event–time pair (e_i, t_i) in r_1 and (e_j, t_j) in r_2, insertion or deletion were used if $e_i \neq e_j$; otherwise, substitution (ie, changing the occurrence time of an event) was used. We set the edit cost to 1 for insertion and deletion operations, $c(\text{Ins}(e))=c(\text{Del}(e))=1$, and the cost of substitution was $c(\text{Sub}(e,t_i,t_j))=0.5*|t_i-t_j|$. Given that we could change sequence r_1 to r_2 via different series of operations, the operation series with the minimum total cost was taken as the edit distance [21].

For instance, for sequence r_1 ={(A, 1), (B, 2), (C, 3), (D, 4)} and r_2 ={(A, 2), (B, 5), (C, 8) }, where (A, 1) indicates that event A occurred in the first day after admission, possible operation series could be Os_1 = {Del(A, 1), Del(B, 2), Del(C, 3), Del(D, 4), Ins(A, 2), Ins(B, 5), Ins(C, 8)}, with a total cost of 7, or Os_2 = {Sub(A, 1, 2), Sub(B, 2, 5), Sub(C, 3, 5), Del(D, 4)}, with a total cost of 5.5; therefore, the second operation series is optimal. We used a dynamic programming algorithm [20] to solve this minimization problem (Multimedia Appendix 1).

The sequence similarity for a pair of event sequences was

where $M(m,n)$ was the edit distance, and m and n were the lengths of sequences r_1 and r_2. Laboratory test items $S_{lab\text{-}edit}$, radiological examinations $S_{rad\text{-}edit}$, and procedures $S_{pro\text{-}edit}$ were represented by 3 individual event sequences.

Time Series Similarity

In the clinical field, a time series can be defined as a consistent, unidirectional change in the value of a biosignal and is, thus, related to the evolution of a patient's status [22]. In this study, a time series s was defined as multiple real values of a laboratory test sorted temporally during a patient's hospitalization. This type of time series often has different lengths because patients with different diseases have different numbers of laboratory test items. In this situation, the traditional Euclidean or cosine distance was not suitable for calculating the similarity between 2 time series. We used dynamic time warping, which has been frequently implemented to assess similarity between time series data [23,24], to calculate the distance between laboratory test time series. The dynamic time warping algorithm applied dynamic programming algorithm, and the cost for each map was defined by the Euclidean distance between 2 time series (Multimedia Appendix 1). By using the dynamic time warping algorithm, we obtained the optimal alignment and the cumulative distance between 2 time series when mapping one time series onto the other [25].

The trend similarity S_{DTW} for s_1 and s_2 was

where $D(s_1,s_2)$ was the final cumulative distance between s_1 and s_2. The minimum and maximum values of all pairwise distances were denoted as d_{min} and d_{max}, respectively.

We also used Haar wavelet decomposition method to assess similarity. The Haar wavelet-based method is highly dependent on time series length; therefore, linear interpolation to ensure time series satisfied length requirements. Using discrete Haar wavelet decomposition, each time series was represented by several Haar wavelet bases (Figure S3 in Multimedia Appendix 1), and the coefficients of these bases, which described main characteristics and changing trends in the time series [26], were used to calculate Haar wavelet-based trend similarity S_{Haar},

where $d(s_1, s_2)$ was the Euclidean distance between 2 groups of coefficients describing s_1 and s_2.

The trend similarities between a laboratory test's multiple test values were calculated using either dynamic time warping or Haar wavelet-based decomposition.

Similarity for Cross-sectional Information

Cross-sectional information comprised demographic characteristics (age, sex, payment type, and marital status), laboratory tests only performed at admission, and free-text reports of chest x-rays and color sonography.

Demographic characteristics were represented as 0 or 1 in vector u based on whether or not the patient was ≥60 years, male, married, and insured (specific medical insurance). To assess demographic feature similarity for patients i and j, we used Jaccard similarity

We calculated Euclidean distance–based similarities for laboratory tests performed only at admission. The feature similarity for these cross-sectional laboratory tests (S_{lab}) was defined as 1 – normalized Euclidean distance, using minimum–maximum normalization.

The free-text reports were in English in the public data set and Chinese in the private data set. For reports written in Chinese, we performed 3 steps to extract features: corpus-of-interest construction, word segmentation, and feature reconstruction (Figure S4 in Multimedia Appendix 1). For reports written in English, we directly identified features of interest (high frequency of occurrence and related to acute myocardial infarction, for example "LVEF," because patients with high LVEF usually have better cardiac function and prognosis). A text feature variable was set to 1 if patients' radiological reports contained this feature and 0 otherwise. Finally, each patient had a set of h features from text such as "左室射血分数正常 (Left ventricular ejection fraction (LVEF) was normal)" in the private data set and "Overall normal LVEF" in the public data set. We used Jaccard similarity to calculate similarity for extracted text features (S_{text}).

Patient Similarity Calculation

The patient similarity score was the weighted sum of feature similarities. We identified dominant features, to which greater weights were assigned, and set weights for the rest of the features to 0. Weights were assigned separately for mortality or readmission risk prediction tasks. The importance of a feature was determined by the predictive performance when using the similarity calculated on this feature to identify nearest neighbors for death or readmission prediction. The greater the performance, the greater the feature importance. Based on the sample set for weight determination, death risk, for example, of an index patient was predicted as the occurring probability of outcomes status among his top *k* nearest neighbors. We selected near neighbors using the similarity of one of the following features in turn: event sequences of laboratory test items, radiological examinations, and procedures; time series of lab tests having multiple testing values; and cross-sectional features.

We identified 3 dominant features, with a majority voting scheme, that had the highest area under receiver operating characteristic curve (AUROC) values. We optimized feature weights w_1, w_2, and w_3, by 0.05 steps, under the constraints $w_1 + w_2 + w_3 = 1$ and $w_1 \geq w_2 \geq w_3 > 0$ (Multimedia Appendix 1).

Predictive Models

Similarity–Based Model Configuration

We built several *k*–nearest neighbor classifiers to predict patients' outcomes based on patient similarity.

We compared predictive performances of *k*–nearest neighbor models built using the sequence similarity alone, the trend similarity alone, and both (Table 1). The subscripts E, D, and H represented the *k*-nearest neighbor model was built by using sequence similarity alone, the dynamic time warping-based trend similarity alone, and Haar decomposition-based trend similarity alone, separately. The subscript ED indicated the *k*-nearest neighbor model was built on both sequence similarity and trend similarity using dynamic time warping, while EH indicated the trend similarity was measured using Haar decomposition.

Table 1. The construction of similarity-based predictive models based on different patient similarities.

Similarity used	*k*–nearest neighbor$_{ED}$	*k*–nearest neighbor$_{EH}$	*k*–nearest neighbor$_{E}$	*k*–nearest neighbor$_{D}$	*k*–nearest neighbor$_{H}$
Sequence similarity					
$S_{\text{lab-edit}}$	Yes	Yes	Yes	No	No
$S_{\text{rad-edit}}$	Yes	Yes	Yes	No	No
$S_{\text{pro-edit}}$	Yes	Yes	Yes	No	No
Trend similarity					
S_{DTW}	Yes	No	No	Yes	No
S_{Haar}	No	Yes	No	No	Yes
Cross-sectional information–based similarity (S_{dem}, S_{lab}, S_{text})	Yes	Yes	Yes	Yes	Yes

Comparison Model Configuration

We compared the predictive performance of each *k*–nearest neighbor model with those of other state-of-the-art predictive models: Euclidean-distance *k*–nearest neighbor, logistic regression, random forest, long short-term memory network, and recurrent neural network models, using either the full set of predictor variables or a set of statistical features, because time series data could not be directly input to Euclidean-distance *k*–nearest neighbor, logistic regression, or random forest models. Cross-sectional information and all flattened time series (padded and concatenated) were input to Euclidean-distance *k*–nearest neighbor, logistic regression, and random forest models, and a set of 6 statistical features for each time series—minimum, maximum, mean, standard deviation, skewness, and time series length—were input to each model with the cross-sectional information. The model with the higher performance for the 2 abovementioned strategies was reported and compared with our similarity-based models (Table S1 in Multimedia Appendix 1).

Model Hyperparameters

We searched for the optimal parameters of models by trial and error. Finally, we set *k*=50 for *k*–nearest neighbor and the number of trees to 200 for the random forest model. For the training of logistic regression, long short-term memory network, and recurrent neural network models, we defined loss functions as cross-entropy with an L2-regulation term. The long short-term memory network and recurrent neural network were trained with an adaptive moment estimation optimizer with a sigmoid activation function. For long short-term memory network and recurrent neural network models, the number of units was set to 100, batch size was chosen as 128, and the maximum number of epochs was set to 30. The leave-one-out method was used to evaluate performances of predictive models, with one patient used as a test sample and the rest used for training in each validation round. This method made full use of the validation set and can be used with an imbalanced data set.

Because we aimed to provide an early warning to allow for timely intervention and treatment adjustment, 3 time points, at admission, Day 7, and at discharge were denoted as the index

time points. All available information at each index time point was used for determining patient similarity and building predictive models. To ensure robustness, we ran the predictive process 100 times independently and averaged the performances. The differences between models' performances were considered statistically significant if model A outperformed model B at least 95 times. AUROC and F1-score were used as the main metrics and we also calculated precision, sensitivity, and specificity.

Data Set and Features

Public Data Set

We used the freely accessible critical care database Medical Information Mart for Intensive Care III (MIMIC-III) [27,28]. The MIMIC-III data set was collected between June 2001 and October 2012 from patients admitted to intensive care units at the Beth Israel Deaconess Medical Center in Boston, Massachusetts. It includes patient health information such as demographic data, vital signs, laboratory test results, medications, procedures, diagnosis codes, as well as clinical notes. In this study, we included all records for patients with acute myocardial infarction.

A total of 3010 patients whose primary diagnosis, confirmed with International Classification of Diseases ninth revision codes 410.01 to 410.91, were enrolled in this study. We extracted data on age at admission, sex, payment type, marital status, 42 laboratory tests (23 discrete time series and 19 cross-sectional items), procedures, and radiology reports (34 text features; Table S2 in Multimedia Appendix 1) during hospitalization.

Private Data Set

Electronic medical record data used in this study were derived from records of inpatients discharged from a tertiary hospital in Beijing, China between 2014 and 2016. Individual hospitalizations were deidentified and maintained as unique records. Overall, 1846 patients whose primary diagnosis confirmed with the International Classification of Disease, tenth revision, codes I21 and I22 were enrolled. Of the laboratory tests, 103 laboratory tests were used as cross-sectional information (at admission). By Day 7, 27 laboratory tests had 2 or more testing values, and the rest were used as cross-sectional information. At discharge, 63 and 40 laboratory test items were treated as time series and cross-sectional information, respectively. For radiological reports, a set of 36 text features (Table S2 in Multimedia Appendix 1) was obtained.

Inclusions and Exclusions

For both data sets, few patients underwent a chest x-ray or a color sonography examination during the first week after admission; therefore, text features were not extracted from radiological reports for further similarity calculation when using information before the first week after admission. The event sequence that comprised radiological examinations was also excluded from sequence similarity calculation because few events occurred at admission. Additionally, a total of 164 patients with a length of stay less than 7 days were excluded from the training sample set when the prediction was made for Day 7. Patients with any length of stay were included in the prediction using patient information during a hospitalization episode. Only 33 and 52 patients in the private data set were readmitted within 30 and 90 days, respectively. Thus, no time requirement was used to identify readmission.

Results

General

Table 2 presents characteristics and main outcomes of the study population.

Table 2. Basic characteristics of acute myocardial infarction patients in MIMIC-III data set and the private data set.

Characteristic	MIMIC-III data set (n=3010), n (%)	Private data set (n=1846), n (%)
Demographic		
Age ≥60 years	2408 (80.0)	1131 (61.3)
Male gender	1855 (61.6)	1343 (72.8)
Married	1583 (52.6)	1815 (98.3)
Medical Insurance		
Urban Employee Basic	N/A[a]	1422 (77.0)
Medicare	2030 (67.4)	N/A
Events during a hospital stay, n (numbers of events per patient)		
Laboratory test	1,044,886 (347)	349,563 (189)
Radiological examination	19,171 (6)	5827 (3)
Procedure	19,630 (7)	13,049 (7)
Outcomes		
Acute myocardial infarction-cause readmission, n (%)	554 (18.4)	100 (5.4)
All-cause in-hospital mortality, n (%)	245 (8.2)	132 (7.2)
Length of hospital stay, day, mean (standard deviation)	10.0 (6.24)	11.4 (5.85)

[a]N/A: not applicable.

Public Data Set

When predicting mortality, all *k*–nearest neighbor models built on patient similarities involving events performed best (*k*–nearest neighbor$_E$: AUROC 0.878; *k*–nearest neighbor$_{EH}$: AUROC 0.882; and *k*–nearest neighbor$_{ED}$: AUROC 0.883) and significantly outperformed all other models (random forest: *P*=.02; all other models: *P*<.001) (Table 3 and Figure 3A). For predicting acute myocardial infarction-cause readmission, *k*–nearest neighbor$_E$, *k*–nearest neighbor$_{EH}$ and *k*–nearest neighbor$_{ED}$ also had the highest AUROC values (Table 3), and the 3 *k*–nearest neighbor models also performed best in mortality and readmission prediction when evaluated with F1-scores. There were no significant differences among *k*–nearest neighbor models involving events for mortality (*k*–nearest neighbor$_E$ and *k*–nearest neighbor$_{EH}$: *P*=.44; *k*–nearest neighbor$_{ED}$ and *k*–nearest neighbor$_E$: *P*=.24; *k*–nearest neighbor$_{EH}$ and *k*–nearest neighbor$_{ED}$: *P*=.41) and readmission predictions (*k*–nearest neighbor$_E$ and *k*–nearest neighbor$_{EH}$: *P*=.84; *k*–nearest neighbor$_{ED}$ and *k*–nearest neighbor$_E$: *P*=.73; *k*–nearest neighbor$_{EH}$ and *k*–nearest neighbor$_{ED}$: *P*=.59) (Figure 3).

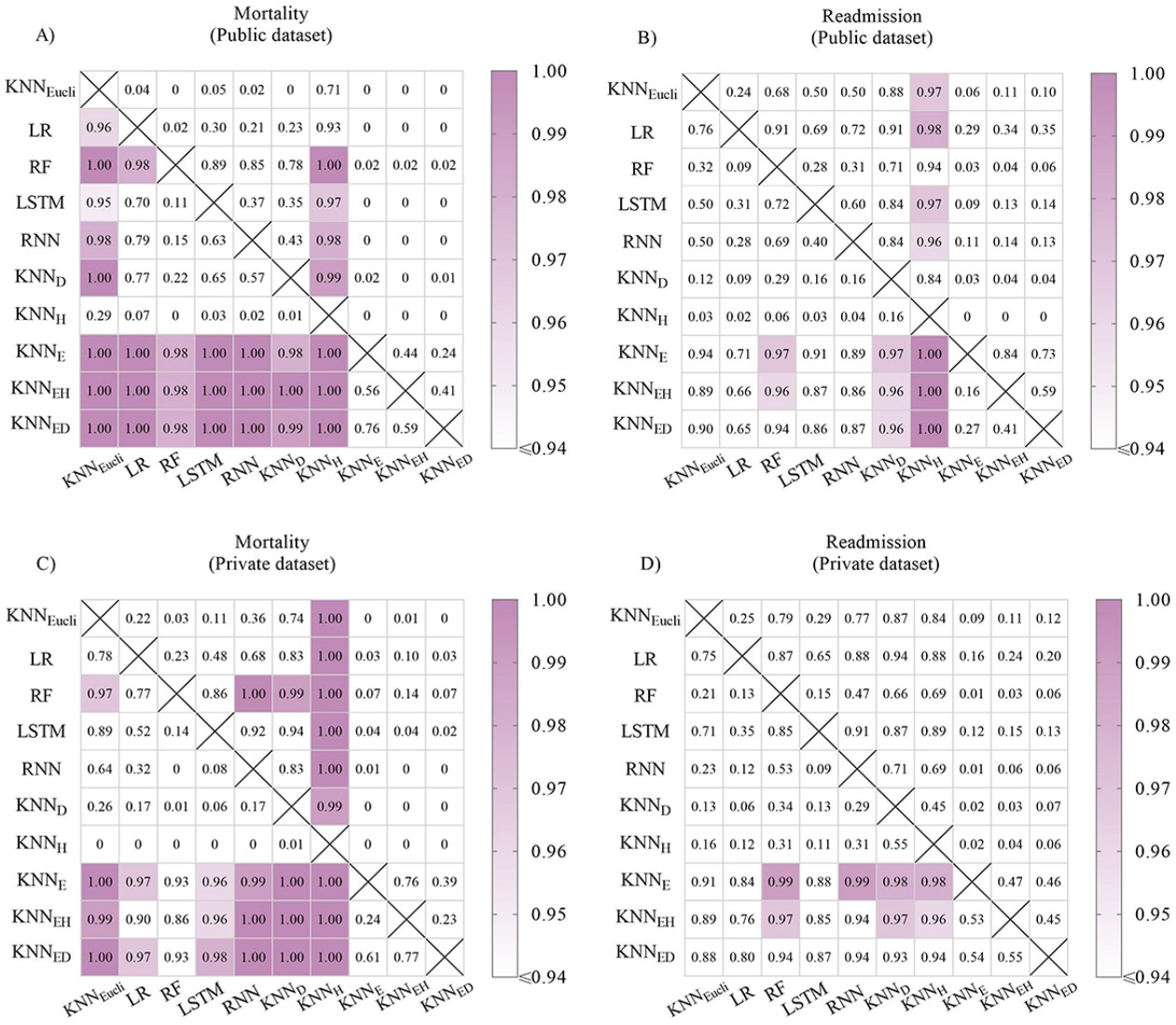
Table 3. The predictive performance of 100 independent rounds of the outcome prediction on the MIMIC-III data set[a].

Model	Mortality		Readmission	
	AUROC[b]	F1-score	AUROC	F1-score
Euclidean distance *k*–nearest neighbor	0.756 (0.022)	0.280 (0.030)	0.592 (0.019)	0.332 (0.019)
Logistic regression	0.796 (0.024)	0.336 (0.037)	0.608 (0.022)	0.347 (0.019)
Random forest	0.834 (0.015)	0.362 (0.033)	0.579 (0.015)	0.327 (0.020)
Long short-term memory network	0.809 (0.022)	0.356 (0.043)	0.595 (0.020)	0.339 (0.017)
Recurrent neural network	0.814 (0.018)	0.338 (0.039)	0.590 (0.018)	0.337 (0.018)
k–nearest neighbor$_D$	0.816 (0.023)	0.373 (0.047)	0.566 (0.022)	0.315 (0.027)
k–nearest neighbor$_H$	0.746 (0.026)	0.295 (0.035)	0.536 (0.026)	0.295 (0.048)
k–nearest neighbor$_E$	0.878 (0.017)	0.386 (0.041)	0.623 (0.019)	0.350 (0.018)
k–nearest neighbor$_{EH}$	0.882 (0.016)	0.401 (0.044)	0.620 (0.018)	0.350 (0.018)
k–nearest neighbor$_{ED}$	0.883 (0.015)	0.406 (0.050)	0.620 (0.019)	0.351 (0.019)

[a]Mean: standard deviation.

[b]AUROC: area under the receiver operating characteristic curve.

Figure 3. Heatmaps showing the pairwise comparisons among models for predicting mortality (A and C) and readmission (B and D) based on the public (A and B) and private (C and D) dataset. Number in each cell is the percent of times that the model in row had a higher performance than the model in column after 100 experiments. The performance is considered significantly higher if the number is greater than or equal to 0.95, and the corresponding cell is highlighted in color. KNN_{Eucli}: Euclidean distance k–nearest neighbor; KNN_D: kNN built on the dynamic time warping (DTW) -based trend similarity (ie, k–nearest neighbor$_D$); KNN_H: kNN built on the Haar-based trend similarity (ie, k–nearest neighbor$_H$); KNN_E: kNN built on the sequence similarity (ie, k–nearest neighbor$_E$); KNN_{EH}: kNN built on the sequence similarity and Haar-based trend similarity (ie, k–nearest neighbor$_{EH}$); KNN_{ED}: kNN built on the sequence similarity and DTW-based trend similarity (ie, k–nearest neighbor$_{ED}$); LR: logistic regression; RF: random forest; RNN: recurrent neutral network; LSTM: long short-term memory.

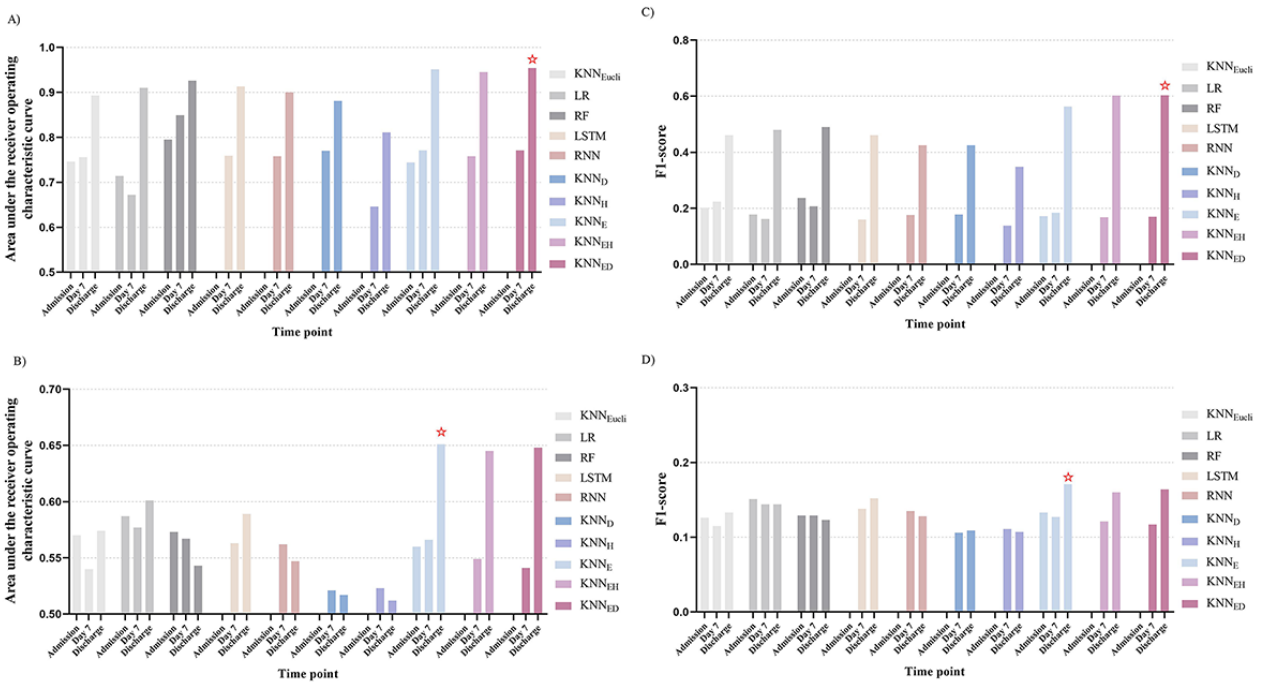


Private Data Set

When predicting mortality, *k*–nearest neighbor$_{ED,}$ which uses both edit distance–based sequence similarity and dynamic time warping–based trend similarity had the best performance (AUROC 0.954; F1-score 0.603) when using all available information from admission to discharge. It significantly outperformed all other state-of-the-art models (Euclidean distance *k*–nearest neighbor: *P*<.001; recurrent neural network: *P*<.001; logistic regression: *P*=.03; long short-term memory network: *P*=.02) except for random forest (at admission: AUROC 0.795; before Day 7: AUROC 0.849; *P*=.07). (Figure 3C and Figure 4A). Predictive performances of all models improved with time points (at admission, Day 7, and at discharge) except for the logistic regression model (Figure 4A).

For readmission prediction, *k*–nearest neighbor$_E$ (AUROC 0.651), *k*–nearest neighbor$_{EH}$ (AUROC 0.645), and *k*–nearest neighbor$_{ED}$ (AUROC 0.648) performed best when using all available information from admission to discharge; however, logistic regression performed best at admission (AUROC 0.589) and before Day 7 (AUROC 0.577) (Figure 4B). The precision, sensitivity, and specificity results of all models are presented in Table S3 (Multimedia Appendix 1).

Figure 4. The predictive performance of all models for predicting inpatient mortality (A and C) and readmission (B and D) in terms of area under the receiver operating characteristic curve (A and B) and F1-score (C and D). Stars (☆) indicate the highest predictive performances. No prediction was made at admission for LSTM, RNN, KNN_H, KNN_D, KNN_{EH}, and KNN_{ED} because of no available temporal information at that time. KNN_{Eucli}: Euclidean distance k–nearest neighbor; KNN_D: kNN built on the dynamic time warping (DTW) -based trend similarity (ie, k–nearest $neighbor_D$); KNN_H: kNN built on the Haar-based trend similarity (ie, k–nearest $neighbor_H$); KNN_E: kNN built on the sequence similarity (ie, k–nearest $neighbor_E$); KNN_{EH}: kNN built on the sequence similarity and Haar-based trend similarity (ie, k–nearest $neighbor_{EH}$); KNN_{ED}: kNN built on the sequence similarity and DTW-based trend similarity (ie, k–nearest $neighbor_{ED}$); LR: logistic regression; RF: random forest; RNN: recurrent neutral network; LSTM: long short-term memory.



Discussion

It is anticipated that predictive modeling based on electronic medical record data will drive personalized medicine and improve health care quality, with many researchers attempting to predict patients' clinical outcomes, such as death [4,6,7,16,22]; quality of care, such as readmissions [4,7,29,30]; resource utilization, such as length of stay [4,6,31], and diagnoses [6,32]. Patient similarity, calculated based on the electronic medical record data, has improved predictive models' performances [9,10].

The longitudinal information in electronic medical record data includes timestamped event sequence and laboratory test time series, which are informative and valuable for outcome predictions due to the rich information on medical behavior and disease progression. However, both types of sequential information are usually heterogeneous, irregular, and uneven, presenting large challenges in data preprocessing, feature extraction, and similarity measurement. Therefore, we used 2 strategies to calculate similarity for timestamped event sequence and laboratory test time series separately. The edit distance, which has been widely used to measure distance in analyzing textual strings [33], biological sequences [34], and patient traces [31], was used to calculate similarity for timestamped event sequences.

For time series, 2 main groups of algorithms for similarity calculation can be identified: the time domain algorithm and the transform-based methods [22]. The former worked directly with the raw time series, while the latter reduced original data dimension for further similarity calculation [22]. We used both a time domain (dynamic time warping) and transform-based (Haar wavelet decomposition) to calculate the trend similarity for time series. Dynamic time warping worked better in trend similarity calculations than Haar wavelet decomposition, based on the results for both data sets. Haar wavelet-based trend similarity methods might not be suitable for time series in electronic medical record system. because more information is lost during dimension reduction than that in dynamic time warping. Our findings that dynamic time warping for time-varying features increased predictive performances were similar to those from a previous study [35]. The most frequently selected features were the procedure-based sequence, the serum creatinine level, and the radiological examination-based sequence. This finding inspired us to shed more light on event sequence and specific clinical variables, which helped in identifying similar patients and improving downstream personalized prediction. Generally, dynamic time warping and the edit distance could be used with sequential information having different lengths and helped overcome the challenge of evaluating sequential similarity for uneven electronic medical record data.

Classical time series processing models, such as recurrent neural network and long short-term memory network, could not use event sequence information, and truncation or 0-padding was inevitable in order to process time series with different lengths. Whereas, *k*–nearest neighbor models based on the proposed patient similarity measurement can make use of 2 types of

sequential information and performed best in outcome prediction in this study. To the best of our knowledge, this is the first study in which 2 types of sequential information have been integrated and applied to patient similarity measurement. Furthermore, the predictive mechanisms of *k*–nearest neighbor models are more interpretable and transparent for clinicians than some black box models such as random forest, recurrent neural network, and long short-term memory network [16]. In general, our models helped improve predictive performance.

Several prior studies evaluated model performances and compared them with other experiments conducted on the MIMIC-III data set. Zhang et al [4] proposed a fusion model leveraging sequential clinical notes, time series, and static information (AUROC 0.871) that outperformed baseline models for mortality prediction. Guo et al [36] constructed a nomogram to predict in-hospital mortality for myocardial infarction patients (AUROC 0.803). Jiang et al [37] used machine learning to predict in-hospital mortality in sepsis survivors (sepsis: AUROC 0.732; nonsepsis: AUROC 0.830). Suresh et al [38] developed a multitask model (AUROC 0.869) for mortality prediction that outperformed global and separate models. Fan et al [39] predicted in-hospital mortality for acute myocardial infarction patients by building several models such as logistic regression, decision tree, extreme gradient boosting, and random forest; among which, the logistic regression model performed best (AUROC 0.870). In this study, the sequential similarity–based model (AUROC 0.883 for the MIMIC-III data set) had better predictive performance for mortality prediction that those mentioned. The model successfully measured the closeness among patients, helped selecting similar study cohort, and assisted building personalized predictive models. Furthermore, we found that sequence similarity was better at identifying nearest neighbors than trend similarity. This finding coincided with the conclusion that patients' clinical traces were informative, and similar patient traces might have similar endpoints [31].

Early detection of endpoints for at-risk patients is key for understanding and improving outcomes [5]. In our study, we selected 3 timepoints during hospitalization: at admission, Day 7, and at discharge. At each timepoint, all available data including sequential information were used to predict the outcomes of patients with acute myocardial infarction. For predicting mortality, the performances of all predictive models, except logistic regression, improved with the 3 timepoints. This finding indicated that sequential data helped improve performances of models. The more sequential information involved, the better the predictive performance. This finding verified our initial assumption that longitudinal information in conjunction with patient similarity measurement would facilitate more accurate outcome prediction.

For predicting unplanned readmission, our model performed best on both data sets when all data, from during the whole hospitalization period, were used. This finding sufficiently indicated that patient similarity could significantly boost the performance of readmission prediction. However, unsatisfactory predictive results for readmission prediction were found in our study and have also been found in other studies [4,6]. The reason might be that the readmission condition was multifactorial and complex, such as related to patient medical insurance, economic conditions, and individual factors, thus, it is challenging to make a prediction [4]. In addition, we noted that the performances of all models for mortality and readmission prediction at admission and at Day 7 were significantly lower than those at discharge, possibly because information that is a long temporal interval from discharge was not useful in outcome prediction.

This study had some limitations. First, trend similarity can also be calculated based on time series in the form of abnormality status. This method would require validation in the future. Second, the patient information used in this study was insufficient. The electrocardiogram captures vital signs for patients with acute myocardial infarction and a type of longitudinal information enabling temporal similarity calculation. However, this information was unavailable for the private data set. Therefore, electrocardiograms should be collected and used for similarity measurement in further study.

In this study, we proposed a complete framework for measuring patient similarity that used both sequential and cross-sectional information. The method successfully evaluated sequential similarity, helped deal with the challenge of similarity calculation for uneven electronic medical record data, and improved the performance of predicting patients' outcomes.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grants 81971707 and 81671786) and Beijing Advanced Innovation Center for Big Data-based Precision Medicine (grant number PXM2021_014226_000026).We are grateful to Dr. Yinjing Hou (Beijing Tongren Hospital, Capital Medical University, Beijing, China) for her clinical advice.

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplementary material.
[DOCX File , 700 KB - jmir_v24i1e30720_app1.docx]

References

1. Qiu H, Yu H, Wang L, Yao Q, Wu S, Yin C, et al. Electronic health record driven prediction for gestational diabetes mellitus in early pregnancy. Sci Rep 2017 Nov 27;7(1):16417-16413 [FREE Full text] [doi: 10.1038/s41598-017-16665-y] [Medline: 29180800]
2. Bahl M, Barzilay R, Yedidia AB, Locascio NJ, Yu L, Lehman CD. High-risk breast lesions: a machine learning model to predict pathologic upgrade and reduce unnecessary surgical excision. Radiology 2018 Mar 17;286(3):810-818. [doi: 10.1148/radiol.2017170549] [Medline: 29039725]
3. Suo Q, Ma F, Yuan Y, Huai M, Zhong W, Gao J, et al. Deep patient similarity learning for personalized healthcare. IEEE Trans Nanobioscience 2018 Jul;17(3):219-227. [doi: 10.1109/tnb.2018.2837622]
4. Zhang D, Yin C, Zeng J, Yuan X, Zhang P. Combining structured and unstructured data for predictive models: a deep learning approach. BMC Med Inform Decis Mak 2020 Oct 29;20(1):280-211 [FREE Full text] [doi: 10.1186/s12911-020-01297-6] [Medline: 33121479]
5. Harutyunyan H, Khachatrian H, Kale DC, Ver Steeg G, Galstyan A. Multitask learning and benchmarking with clinical time series data. Sci Data 2019 Jun 17;6(1):96-18 [FREE Full text] [doi: 10.1038/s41597-019-0103-9] [Medline: 31209213]
6. Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ Digit Med 2018 May 8;1(1):18-10 [FREE Full text] [doi: 10.1038/s41746-018-0029-1] [Medline: 31304302]
7. Chu J, Dong W, Huang Z. Endpoint prediction of heart failure using electronic health records. J Biomed Inform 2020 Sep;109:103518 [FREE Full text] [doi: 10.1016/j.jbi.2020.103518] [Medline: 32721582]
8. Esteban C, Staeck O, Baier S, Yang Y, Tresp V. Predicting clinical events by combining static and dynamic information using recurrent neural networks. 2016 Nov 8 Presented at: IEEE International Conference on Health care Informatics; October 4-7; Chicago, Illinois, USA p. 93-101. [doi: 10.1109/ichi.2016.16]
9. Parimbelli E, Marini S, Sacchi L, Bellazzi R. Patient similarity for precision medicine: a systematic review. J Biomed Inform 2018 Jul 1;83:87-96 [FREE Full text] [doi: 10.1016/j.jbi.2018.06.001] [Medline: 29864490]
10. Sharafoddini A, Dubin JA, Lee J. Patient similarity in prediction models based on health data: a scoping review. JMIR Med Inform 2017 Mar 03;5(1):e7 [FREE Full text] [doi: 10.2196/medinform.6730] [Medline: 28258046]
11. Wang N, Huang Y, Liu H, Zhang Z, Wei L, Fei X, et al. Study on the semi-supervised learning-based patient similarity from heterogeneous electronic medical records. BMC Med Inform Decis Mak 2021 Jul 30;21(Suppl 2):58-13 [FREE Full text] [doi: 10.1186/s12911-021-01432-x] [Medline: 34330261]
12. Wang N, Huang Y, Liu H, Fei X, Wei L, Zhao X, et al. Measurement and application of patient similarity in personalized predictive modeling based on electronic medical records. Biomed Eng Online 2019 Oct 11;18(1):98-15 [FREE Full text] [doi: 10.1186/s12938-019-0718-2] [Medline: 31601207]
13. Li L, Cheng W, Glicksberg BS, Gottesman O, Tamler R, Chen R, et al. Identification of type 2 diabetes subgroups through topological analysis of patient similarity. Sci Transl Med 2015 Oct 28;7(311):311ra174-311ra115 [FREE Full text] [doi: 10.1126/scitranslmed.aaa9364] [Medline: 26511511]
14. Wang F, Hu J, Sun J. Medical prognosis based on patient similarity and expert feedback. In: Proceedings of the 21st International Conference on Pattern Recognition. 2013 Feb 14 Presented at: 21st International Conference on Pattern Recognition; November 11-15; Tsukuba, Japan p. 1799-1802.
15. Wu J, Roy J, Stewart W. Prediction modeling using EHR data: challenges, strategies, and a comparison of machine learning approaches. Med Care 2010 Jun;48(6 Suppl):S106-S113. [doi: 10.1097/MLR.0b013e3181de9e17] [Medline: 20473190]
16. Lee J, Maslove DM, Dubin JA. Personalized mortality prediction driven by electronic medical data and a patient similarity metric. PLoS One 2015 May 15;10(5):e0127428 [FREE Full text] [doi: 10.1371/journal.pone.0127428] [Medline: 25978419]
17. Ng K, Sun J, Hu J, Wang F. Personalized predictive modeling and risk factor identification using patient similarity. AMIA Jt Summits Transl Sci Proc 2015;2015:132-136 [FREE Full text] [Medline: 26306255]
18. Sun J, Sow D, Hu J, Ebadollahi S. Localized supervised metric learning on temporal physiological data. In: Proceedings of the 20th International Conference on Pattern Recognition. 2010 Oct 7 Presented at: 20th International Conference on Pattern Recognition; August 23-26; Istanbul, Turkey p. 4149-4152. [doi: 10.1109/icpr.2010.1009]
19. Chang J, Liu X, Sun Y. Mortality due to acute myocardial infarction in China from 1987 to 2014: secular trends and age-period-cohort effects. Int J Cardiol 2017 Jan 15;227:229-238. [doi: 10.1016/j.ijcard.2016.11.130] [Medline: 27839815]
20. Mannila H, Moen P. Similarity between event types in sequences. 1999 Presented at: The First International Conference on Data Warehousing and Knowledge Discovery; September 1; Florence, Italy p. 271-280. [doi: 10.1007/3-540-48298-9_29]
21. Moen P. Attribute, event sequence, and event type similarity notions for data mining. University of Helsinki. URL: https://www.cs.helsinki.fi/u/ronkaine/phdthesis.pdf [accessed 2021-12-14]
22. Henriques J, Carvalho P, Paredes S, Rocha T, Habetha J, Antunes M, et al. Prediction of heart failure decompensation events by trend analysis of telemonitoring data. IEEE J Biomed Health Inform 2015 Sep;19(5):1757-1769. [doi: 10.1109/jbhi.2014.2358715]
23. Gold O, Sharir M. Dynamic time warping and geometric edit distance. ACM Trans Algorithms 2018 Oct 13;14(4):1-17. [doi: 10.1145/3230734]
24. Allam A, Dittberner M, Sintsova A, Brodbeck D, Krauthammer M. Patient similarity analysis with longitudinal health data. ArXiv Preprint posted online on May 14, 2020. [FREE Full text]

25. Giorgino T. Computing and visualizing dynamic time warping alignments in R: the dtw package. J Stat Soft 2009;31(7):1-25. [doi: 10.18637/jss.v031.i07]
26. Panahiazar M, Taslimitehrani V, Pereira N, Pathak J. Using EHRs for heart failure therapy recommendation using multidimensional patient similarity analytics. Stud Health Technol Inform 2015;210:369-373 [FREE Full text] [Medline: 25991168]
27. MIMIC-III. Medical Information Mart for Intensive Care. URL: https://mimic.physionet.org/ [accessed 2020-06-20]
28. Johnson AE, Pollard TJ, Shen L, Lehman LH, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci Data 2016 May 24;3:160035 [FREE Full text] [doi: 10.1038/sdata.2016.35] [Medline: 27219127]
29. Shadmi E, Flaks-Manov N, Hoshen M, Goldman O, Bitterman H, Balicer RD. Predicting 30-day readmissions with preadmission electronic health record data. Med Care 2015 Mar;53(3):283-289. [doi: 10.1097/MLR.0000000000000315] [Medline: 25634089]
30. Nguyen P, Tran T, Wickramasinghe N, Venkatesh S. Deepr: a convolutional net for medical records. IEEE J Biomed Health Inform 2017 Jan;21(1):22-30. [doi: 10.1109/jbhi.2016.2633963]
31. Huang Z, Juarez JM, Duan H, Li H. Length of stay prediction for clinical treatment process using temporal similarity. Expert Syst Appl 2013 Nov;40(16):6330-6339. [doi: 10.1016/j.eswa.2013.05.066]
32. Gottlieb A, Stein GY, Ruppin E, Altman RB, Sharan R. A method for inferring medical diagnoses from patient similarities. BMC Med 2013 Sep 02;11(1):194-199 [FREE Full text] [doi: 10.1186/1741-7015-11-194] [Medline: 24004670]
33. Lu W, Du X, Hadjieleftheriou M, Ooi BC. Efficiently supporting edit distance based string similarity search using B-Trees. IEEE Trans. Knowl. Data Eng 2014 Dec 1;26(12):2983-2996. [doi: 10.1109/tkde.2014.2309131]
34. Jiang T, Lin G, Ma B, Zhang K. A general edit distance between RNA structures. J Comput Biol 2002 Apr;9(2):371-388. [doi: 10.1089/10665270252935511] [Medline: 12015887]
35. Hoogendoorn M, Hassouni A, Mok K, Ghassemi M, Szolovits P. Prediction using patient comparison vs. modeling: a case study for mortality prediction. 2016 Presented at: 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC); August 16-20; Orlando, Florida p. 2464-2467. [doi: 10.1109/embc.2016.7591229]
36. Guo Q, Wu M, Li H, Ouyang H, Sun R, Wang J, et al. Development and validation of a prognostic nomogram for myocardial infarction patients in intensive care units: a retrospective cohort study. BMJ Open 2020 Dec 17;10(12):e040291 [FREE Full text] [doi: 10.1136/bmjopen-2020-040291] [Medline: 33334835]
37. Jiang Z, Bo L, Xu Z, Song Y, Wang J, Wen P, et al. An explainable machine learning algorithm for risk factor analysis of in-hospital mortality in sepsis survivors with ICU readmission. Comput Methods Programs Biomed 2021 Jun;204:106040. [doi: 10.1016/j.cmpb.2021.106040] [Medline: 33780889]
38. Suresh H, Gong J, Guttag J. Learning tasks for multitask learning. In: Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018 Presented at: 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining; August 19-23; London, United Kingdom p. 802-810. [doi: 10.1145/3219819.3219930]
39. Fan Y, Zhao Y, Li P, Liu X, Jia L, Li K, et al. [Analysis of diseases distribution in Medical Information Mart for Intensive Care III database]. Zhonghua Wei Zhong Bing Ji Jiu Yi Xue 2018 Jun;30(6):531-537. [doi: 10.3760/cma.j.issn.2095-4352.2018.06.006] [Medline: 30009726]

Abbreviations

AUROC: area under the receiver operating characteristic curve
MIMIC-III: Medical Information Mart for Intensive Care III

Edited by R Kukafka; submitted 26.05.21; peer-reviewed by J Lei, J Chen, Y Li; comments to author 31.08.21; revised version received 08.10.21; accepted 08.11.21; published 06.01.22.

Please cite as:
Wang N, Wang M, Zhou Y, Liu H, Wei L, Fei X, Chen H
Sequential Data–Based Patient Similarity Framework for Patient Outcome Prediction: Algorithm Development
J Med Internet Res 2022;24(1):e30720
URL: https://www.jmir.org/2022/1/e30720
doi:10.2196/30720
PMID:34989682

Letter to the Editor

Using Social Media in Health Care Research Should Proceed With Caution. Comment on “The Use of Social Media for Health Research Purposes: Scoping Review”

Abdias Girardi[1], BS; Nikhi Paul Singh[1], BS; Carter Joseph Boyd[2], MD, MBA

[1]Heersink School of Medicine, University of Alabama at Birmingham, Birmingham, AL, United States

[2]Hansjörg Wyss Department of Plastic Surgery, New York University Langone Health, New York, NY, United States

Corresponding Author:
Nikhi Paul Singh, BS
Heersink School of Medicine
University of Alabama at Birmingham
1670 University Blvd
Birmingham, AL, 35233
United States
Phone: 1 2055633208
Email: nsingh@uab.edu

Related Article:

Comment on: https://www.jmir.org/2021/5/e25736

Abstract

(J Med Internet Res 2022;24(1):e35286) doi:10.2196/35286

KEYWORDS

public health; epidemiology; research; health; medical; social networking; infodemiology; eHealth; text mining; medical education; social media; information technology; health care; HIPAA; education

We thank Bour and colleagues for their recent publication regarding the use of social media for health research purposes [1]. Many people use social media to gain knowledge about medicine, and as such, it is important to characterize the reliability of research health topics on these platforms.

A major advantage of social media is its use as an effective vehicle for large-scale dissemination of information, as seen throughout the COVID-19 pandemic. In addition to facilitating health surveillance and disseminating public health information, social media has been helpful in providing social support and behavioral counseling for people in need of such services [2]. Lack of transportation is often a barrier to health care access, and although not every medical condition can be addressed online, access to social media may assist in health education.

Social media has limitations though. Ethical issues concerning consent, privacy, and confidentiality of users are commonly encountered. Social media is considered to be public, and user consent is not provided while collecting social media data. Health care professionals worry about breaching patient confidentiality and facing consequences under the federal Health Insurance Portability and Accountability Act (HIPAA) and state privacy laws. Although it is acceptable to share deidentified patient information, a study of medical blogs proved that keeping patient information deidentified might not be as easy as it seems—individual patients were described in 42% of the 271 samples studied [3]. In this cohort, 17% of cases were found to include enough information for patients to identify themselves or their providers [3]. Moreover, social media users have a skewed distribution toward adolescents and young adults. Thus, the representativeness of the sample may be misleading, resulting in biased findings and preventing generalization to the entire population [1].

While social media has the potential to be a useful tool in health research, both factual and false information can be posted online. Currently, no verification system exists for health information on these platforms. Users can be exposed to dangerous and fake content posted by detractors and chatbots on social media. A recent study from the Massachusetts Institute of Technology demonstrated that false claims are 70% more likely to be

retweeted on Twitter than the truth [4]. This represents a serious threat to public health, as misinformation can be readily spread.

Lack of regulation in social media poses a potential risk to patient safety and public well-being. The American Medical Association (AMA) *Journal of Ethics* provides guidance for physicians navigating these platforms and emphasizes the convection of truthful information and confrontation with misleading or false information [5]. The use of a systematic approach and adherence to the AMA's guidelines may prove useful, preventing physicians from accidentally sharing patient information or misinformation. Social media holds great promise in medicine, but caution should be taken.

Conflicts of Interest

None declared.

Editorial Notice

The corresponding author of *"The Use of Social Media for Health Research Purposes: Scoping Review"* declined to respond to this letter.

References

1. Bour C, Ahne A, Schmitz S, Perchoux C, Dessenne C, Fagherazzi G. The Use of Social Media for Health Research Purposes: Scoping Review. J Med Internet Res 2021 May 27;23(5):e25736 [FREE Full text] [doi: 10.2196/25736] [Medline: 34042593]
2. Abroms LC. Public Health in the Era of Social Media. Am J Public Health 2019 Feb;109(S2):S130-S131. [doi: 10.2105/AJPH.2018.304947] [Medline: 30785795]
3. Ventola L. Social media and health care professionals: benefits, risks, and best practices. P T 2014 Jul;39(7):491-520 [FREE Full text] [Medline: 25083128]
4. Brown S. MIT Sloan research about social media, misinformation, and elections. MIT Management Sloan School. 2020 Oct 5. URL: https://mitsloan.mit.edu/ideas-made-to-matter/mit-sloan-research-about-social-media-misinformation-and-elections [accessed 2022-01-26]
5. Wu J, McCormick JB. Why Health Professionals Should Speak Out Against False Beliefs on the Internet. AMA J Ethics 2018 Nov 01;20(11):E1052-E1058 [FREE Full text] [doi: 10.1001/amajethics.2018.1052] [Medline: 30499434]

Abbreviations

AMA: American Medical Association
HIPAA: Health Insurance Portability and Accountability Act

Edited by T Leung; submitted 29.11.21; this is a non–peer-reviewed article;accepted 24.01.22; published 28.01.22.

Please cite as:
Girardi A, Singh NP, Boyd CJ
Using Social Media in Health Care Research Should Proceed With Caution. Comment on "The Use of Social Media for Health Research Purposes: Scoping Review"
J Med Internet Res 2022;24(1):e35286
URL: https://www.jmir.org/2022/1/e35286
doi:10.2196/35286
PMID:35089149

Corrigenda and Addenda

Correction: Computer-Aided Diagnosis of Gastrointestinal Ulcer and Hemorrhage Using Wireless Capsule Endoscopy: Systematic Review and Diagnostic Test Accuracy Meta-analysis

Chang Seok Bang[1,2,3,4], MD, PhD; Jae Jun Lee[3,4,5], MD; Gwang Ho Baik[1,2], MD

[1]Department of Internal Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea

[2]Institute for Liver and Digestive Diseases, Hallym University, Chuncheon, Republic of Korea

[3]Institute of New Frontier Research, Hallym University College of Medicine, Chuncheon, Republic of Korea

[4]Division of Big Data and Artificial Intelligence, Chuncheon Sacred Heart Hospital, Chuncheon, Republic of Korea

[5]Department of Anesthesiology and Pain Medicine, Hallym University College of Medicine, Chuncheon, Republic of Korea

Corresponding Author:
Chang Seok Bang, MD, PhD
Department of Internal Medicine
Hallym University College of Medicine
77 Sakju-ro
Chuncheon, 24253
Republic of Korea
Phone: 82 33 240 5821
Fax: 82 33 241 8064
Email: csbang@hallym.ac.kr

Related Article:

Correction of: https://www.jmir.org/2021/12/e33267

(J Med Internet Res 2022;24(1):e36170) doi:10.2196/36170

In "Computer-Aided Diagnosis of Gastrointestinal Ulcer and Hemorrhage Using Wireless Capsule Endoscopy: Systematic Review and Diagnostic Test Accuracy Meta-analysis" (J Med Internet Res 2021;23(12):e33267), one error was noted.

In the originally published paper, the Acknowledgments section appeared as follows:

> *This work was supported by the Technology development Program (#S2931703) funded by the Ministry of SMEs and Startups (Korea).*

In the corrected version of the paper, the Acknowledgments section has been revised as follows:

> *This work was supported by the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (grant number 2020-0-01604).*

The correction will appear in the online version of the paper on the JMIR Publications website on January 11, 2022, together with the publication of this correction notice. Because this was made after submission to PubMed, PubMed Central, and other full-text repositories, the corrected article has also been resubmitted to those repositories.

Submitted 04.01.22; this is a non–peer-reviewed article;accepted 04.01.22; published 11.01.22.

Please cite as:
Bang CS, Lee JJ, Baik GH
Correction: Computer-Aided Diagnosis of Gastrointestinal Ulcer and Hemorrhage Using Wireless Capsule Endoscopy: Systematic Review and Diagnostic Test Accuracy Meta-analysis
J Med Internet Res 2022;24(1):e36170
URL: https://www.jmir.org/2022/1/e36170
doi:10.2196/36170
PMID:35015660

Original Paper

Direct and Indirect Associations of Media Use With COVID-19 Vaccine Hesitancy in South Korea: Cross-sectional Web-Based Survey

Minjung Lee[1,2], PhD; Myoungsoon You[1,3], PhD

[1]Institute of Health and Environment, Seoul National University, Seoul, Republic of Korea

[2]Office of Dental Education, School of Dentistry, Seoul National University, Seoul, Republic of Korea

[3]Department of Public Health Sciences, Graduate School of Public Health, Seoul National University, Seoul, Republic of Korea

Corresponding Author:
Myoungsoon You, PhD
Department of Public Health Sciences
Graduate School of Public Health
Seoul National University
Gwanak-ro 1
Gwanak-gu
Seoul, 08826
Republic of Korea
Phone: 82 02 880 2773
Email: msyou@snu.ac.kr

Abstract

Background: The battle against the 2019 novel coronavirus (COVID-19) has not concluded. Despite the availability of vaccines, the high prevalence of vaccine hesitancy represents a significant challenge to public health, and raising vaccine acceptance among the public is critical. Although media has become an increasingly popular source of COVID-19 vaccine-related information, the question of whether and how media use is related to the public's vaccine hesitancy warrants exploration.

Objective: This study aimed to (1) examine the level of COVID-19 vaccine hesitancy, (2) identify factors associated with COVID-19 vaccine hesitancy, and (3) explore the direct and indirect relationship between media use and vaccine hesitancy through psychological factors.

Methods: A month before COVID-19 vaccination was initiated in South Korea, we conducted a cross-sectional web-based survey over 6 days (January 20-25, 2021). This study included 1016 participants, and a logit model for regression analyzed associations between sociodemographic factors, health-related factors, psychological factors, and media use toward one's COVID-19 vaccine hesitancy. Additionally, we conducted a path analysis to examine the indirect effects of media use on vaccine hesitancy by using psychological factors (ie, perceived risk of COVID-19 infection, perceived benefits, and perceived barriers of COVID-19 vaccination).

Results: Among the participants (N=1016), 53.3% (541/1016) hesitated to take the COVID-19 vaccine, while 46.7% (475/1016) agreed to accept the vaccine. Of the sociodemographic factors, female gender (odds ratio [OR] 1.967, 95% CI 1.36-2.86; *P*<.001), age in 50s (OR 0.47, 95% CI 0.23-0.96; *P*=.004), and age over 60s (OR 0.49, 95% CI 0.24-0.99; *P*=.04) were significant individual predictors of COVID-19 vaccine hesitancy. Perceived susceptibility of infection (OR 0.69, 95% CI 0.52-0.91; *P*=.01) and perceived benefits of vaccination (OR 0.69, 95% CI 0.52-0.91; *P*=.01) were associated with lower vaccine hesitancy. Perceived barriers of vaccination (OR 1.63, 95% CI 1.29-2.07; *P*<.001) and lower trust in government (OR 0.72, 95% CI 0.53-0.98; *P*=.04) were related to vaccine hesitancy. The use of offline and online media as sources for the perceived benefits of vaccination was associated with vaccine hesitancy, resulting in lower vaccine hesitancy. Moreover, perceived susceptibility of the disease and perceived barriers of vaccination mediated the association between social media use and vaccine hesitancy.

Conclusions: Our findings revealed a considerable level of COVID-19 vaccine hesitancy in South Korea. Gender-based and generation-based public health policies and communication are recommended. Efforts to lower the perceived risk of vaccine side effects and heighten perceived benefits of the vaccine are required. Although the use of media has a positive and negative effect on the population's vaccine hesitancy, efforts should be made to disseminate reliable and timely information on media while confronting misinformation or disinformation for successive implementation of vaccine programs during pandemics.

(J Med Internet Res 2022;24(1):e32329) doi:10.2196/32329

KEYWORDS

COVID-19; coronavirus; vaccination; vaccine hesitancy; media use; social media; public health; pandemic; epidemiology; online information; health information

Introduction

Although the 2019 novel coronavirus (COVID-19) continues to spread worldwide and the public health emergency continues, the battle to overcome COVID-19 remains active. The development of effective vaccines has been highly anticipated, and several vaccines are now available; however, the timeliness of vaccine development and availability are not the only obstacles to overcome from a public health perspective. Raising vaccine acceptance and uptake among the public is essential to elevate public health emergency preparedness, which refers to the level of readiness for public health systems, communities, and individuals to prevent, respond to, and recover from public health emergencies [1]. A sizeable proportion of the population must be vaccinated to reach herd immunity and prevent the continued spread of the virus, and the goal of the South Korean government is to completely vaccinate 70% of the population [2,3]. According to a global survey collected in June 2020, which asked whether they would be vaccinated if the COVID-19 vaccine is proven safe, effective, and if available, 79.79% of the South Korean respondents responded positively [4]. However, there is evidence that the acceptance of the vaccine is declining [5]. A systematic review compared trends in vaccination receptivity over time, and a decrement of vaccine acceptance from >70% (March 2020) to <50% (October 2020) was observed [6].

Despite the availability of vaccination services, the phenomenon of delayed vaccine acceptance or refusal is referred to as vaccine hesitancy [7-9]. Numerous studies have attempted to define and categorize vaccine hesitancy and commonly propose that attitudes toward vaccination exist on a continuum of no demand to high demand and from accepting all vaccines to accepting no vaccines. Generally, vaccine-hesitant individuals are a heterogeneous group in the middle of this continuum [10]. Vaccine-hesitant individuals may refuse some vaccines but agree to others; they may delay or accept vaccines according to the recommended schedule but be unsure in doing so [11,12]. The Strategic Advisory Group of Experts Working Group on Vaccine Hesitancy proposed the determinants of vaccine hesitancy, which are categorized as follows: (1) contextual influences (eg, communication and media environment, historical influences, politics and policies), (2) individual and group differences (eg, personal, family and community members' experience with vaccination, beliefs, attitudes about health and prevention, trust in health system, knowledge, perceived risk and benefits), and (3) vaccine-specific issues (eg, risk and benefit based on epidemiological and scientific evidence, introduction of a new vaccine, costs) [10].

Vaccine hesitancy is complex and context-specific, varying across time, place, and vaccines. Therefore, hesitation of COVID-19 vaccination should be understood in that context. Epidemiologically, COVID-19 is a highly infectious disease with a basic reproduction number of between 2 and 3, and there is a sharp increase in the number of confirmed cases worldwide [13]. COVID-19 vaccines were developed very shortly within a year, although it takes an average of 10.7 years to develop a new vaccine [14]. Further, a well-known aspect of the COVID-19 vaccination is that although the side effects occur less frequently, they can occur [15-17]. In this context, previous studies revealed the influence of individual and group differences on vaccine hesitancy such as perceived risk of COVID-19 infection [18,19], confidence in the capacity of health services to respond to the COVID-19 pandemic [20], and trust in authorities [21,22]. Vaccine-specific issues such as confidence in the efficacy of the vaccines [23-26], fear of side effects [19,25], and high conspiracy beliefs around the vaccines [27] were also revealed to predict COVID-19 vaccine hesitancy.

Along with the factors mentioned above, scholars paid attention to the effects of using various media (eg, offline media, online media, social media) as vaccine-related information sources on vaccine hesitancy. Vaccine-related health information has shaped perceptions, attitudes, and emotions related to the vaccine. Using offline media such as listening to the radio and reading the newspaper frequently was associated with increased vaccination odds [28,29]. Health information seekers who used newspaper articles as a vaccine-information source were more likely to perceive a vaccine as effective, being more likely to accept the influenza vaccine [30]. Refusing or delaying vaccinations of their children by parents was associated with using online media information sources [31,32]. Respondents among Medicare beneficiaries in the United States who relied on webpages for vaccine-related information were likely to be hesitant about COVID-19 vaccine uptake [33]. Social media platforms have become an increasingly popular source of health information, and growing interest has emerged in the role of social media in public health promotion. In particular, 2-way communication between health authorities and the public via social media is possible, and real-time exchange of health information among families and friends during a pandemic is possible [34-36]. The internet is widely used by authorities to inform the public about the latest news, disseminate public health knowledge, refute rumors, and facilitate effective coordination of medical, public, and pharmaceutical resources [37]. However, misinformation and rumors regarding COVID-19 vaccines have also emerged on social media platforms widely [38]. Engagement with vaccine-related information on social media was related to lower perceived vaccine efficacy [30], higher belief that vaccines are unsafe [39], higher conspiracy beliefs regarding the COVID-19 pandemic [40], and lower vaccination rates [34].

There is limited evidence about the acceptance or hesitancy of the COVID-19 vaccine in practice, and the influencing factors of vaccine hesitancy and whether and how media use can influence the public's vaccine hesitancy warrants further

exploration. Therefore, this study conducted a survey a month before the start of vaccination and addressed the level of vaccine hesitancy and investigated factors related to vaccine hesitancy along with which populations to prioritize in COVID-19 vaccination interventions. Moreover, we examined how media use interacts with psychological factors for vaccine hesitancy. Specifically, this study aimed to (1) examine the level of COVID-19 vaccine hesitancy, (2) quantify and test the relationships between sociodemographic, health-related, psychological factors, media use, and COVID-19 vaccine hesitancy, and (3) examine how psychological factors interplay with the media use on vaccine hesitancy. Implications for developing and implementing evidence-based interventions and policies to raise vaccine acceptance and uptake are also discussed in this paper.

Methods

Study Design and Sampling

We conducted a web-based cross-sectional survey on January 20, 2021—a month before vaccination was initiated in South Korea. The questionnaire, which consisted of 83 questions, was developed to (1) evaluate the public's hesitancy of the COVID-19 vaccine and (2) assess the association with health-related factors, psychological factors, and media use by using an anonymous web-based questionnaire. The survey was conducted via a web-based platform from a research company called Korea Research. The company recruited participants by sending survey invitations containing general information about the survey, such as its aim and consent statement via email or text messages, and then registered survey panel members who met the inclusion criteria. The inclusion criteria were as follows: (1) 18 years or older, (2) a resident in South Korea, and (3) a Korean speaker. The company sampled the participants by age, sex, and geographic region–based proportional and quota sampling process. The respondents provided electronic informed consent that appeared on the first page of the survey, and the company protected the confidentiality of the anonymous respondents. Over 1033 participants completed the surveys, and 1016 were included in the analysis after excluding incomplete responses. This study was reviewed and approved by the Institutional Review Board at Seoul National University (IRB 2101/003-005), Seoul, South Korea. All participants provided their informed consent upon enrollment. The data collection took place over 5 days (January 20-25, 2021), a year after the Korea Centers for Disease Control and Prevention confirmed the first case at the early stage of the epidemic (January 20, 2020).

Measurements

Dependent Variables

A 5-point scale questionnaire measured the intention to be vaccinated for COVID-19. Participants were asked, "If a vaccine for coronavirus (COVID-19) becomes available, would you want to receive it?" and provided response options "Definitely not," "Probably would not," "half and half," "Probably would," and "Definitely want to receive it." The COVID-19 vaccination intention response options of "Definitely not," "Probably would not," and "half and half" were coded as "vaccine hesitancy=1," and the options "probably would" and "Definitely want to receive it" were coded as "vaccine acceptance=0" to create dichotomous "hesitancy" versus "acceptance" variable.

Independent Variables

Sociodemographic Factors

Sociodemographic factors included gender (1=male, 2=female), age, family size (ie, living alone, more than 2 persons), the presence of children at home who attend school (more than one=1, none=0), marital status (ie, married, single, divorced, bereaved), and the participants' residence (urban=1, rural=2). We also assessed education level (1=middle school or below, 2=high school graduate, 3=college and above) and monthly household income in South Korean won (1000 won=US $0.87; 1=<2 million won, 2=2 million to 3.99 million won, 3=4 million to 5.99 million won, and 4=6 million to 7.99 million won, 5=≥8 million won).

Health-Related Factors

Health-related factors included seasonal influenza vaccination history, presence of underlying disease, subjective health, and previous COVID-19 diagnosis for the participants. For seasonal influenza vaccination history, participants were asked, "Have you been vaccinated against the seasonal influenza flu in the last 5 years?" Responses included "every year," "more than once," "maybe once," "never," and "don't know." We grouped the participants as having seasonal influenza vaccination history ("every year," "more than once," and "maybe once") or not ("never" and "don't know"). Subjective health status (poor=1, moderate=2, good=3) was investigated to assess health-related factors. We also investigated the presence of underlying disease by asking participants to indicate all diagnosed underlying diseases (eg, hypertension, dyslipidemia, diabetes, chronic cardiac disease, asthma, cancer). We grouped the participants as being with or without diagnoses of one or more underlying diseases.

Media Used to Obtain COVID-19 Vaccination Information

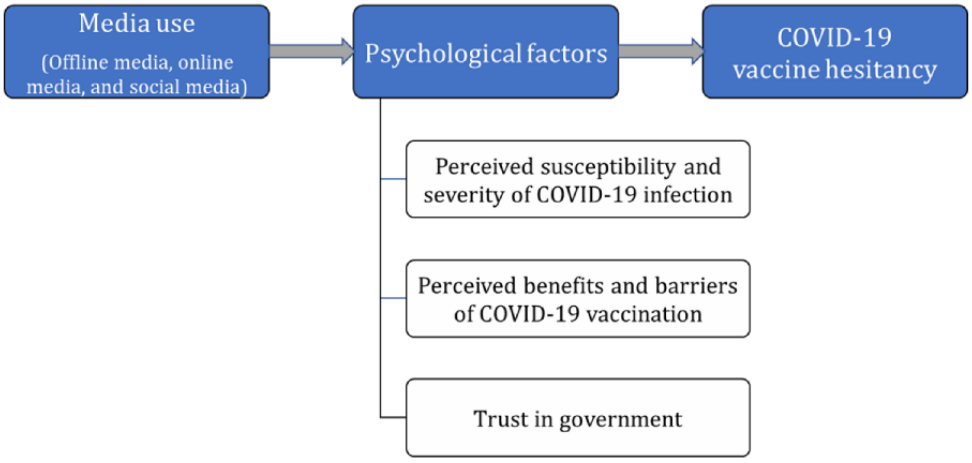
We included the following question to assess participants' media use to obtain vaccine-related health information via various information sources such as offline media (eg, television, radio, newspapers), online media (internet news sites, news portals), and social media (Twitter, Facebook, Kakao Talk, YouTube, blogs, communities). We used a 4-point rating scale (1=not at all to 4=always) to ask the following question: "How often do you use the following information source to seek information about COVID-19 vaccine?"

Psychological Factors Related to COVID-19 Vaccination

Questions to determine the psychological factors that could influence COVID-19 vaccination were adapted from the Health Belief Model and included perceived risk of COVID-19 infection, fear of COVID-19 infection, perceived benefits of COVID-19 vaccination, perceived barriers of vaccination, and trust in the government (Figure 1). Perceived risk of COVID-19 infection comprised 2 components: (1) perceived susceptibility, signifying an individual's beliefs about their possibility of infection, and (2) perceived severity, signifying the seriousness

of infection [25]. Participants were asked, "What do you think is the possibility of COVID-19 infection?" and "What do you think will be the severity if COVID-19 infects you?" Responses were rated on a 5-point rating scale, with "1=very low, 3=neither low nor high, and 5=very high." Perceived benefits of COVID-19 vaccination were measured by 2 items measuring perceived chances of gaining specific benefits by COVID-19 vaccination: (1) self-protection of my health and (2) proved efficacy on preventing COVID-19 infection (1=extremely low to 5=extremely high; Cronbach alpha of .86). Two items addressed the perceived barriers of COVID-19 and included perceived chances of experiencing barriers against COVID-19 vaccination, such as (1) COVID-19 infection caused by vaccination and (2) concern about side effects of vaccination (1=extremely low to 5=extremely high; Cronbach alpha of .61). We also investigated participants' trust in government by asking, "To what extent do you currently trust the government which respond to infectious diseases?" Responses were collected using a 5-point scale, with "1=extremely low to 5=extremely high."

Figure 1. Framework of this study.



Statistical Analysis

All quantitative variables were reported in numbers, proportions, means, and standard deviations. The responses to the COVID-19 vaccination acceptance questions were categorized into binary groups: (1) those who would accept COVID-19 vaccination (acceptance group) and (2) those who hesitated to uptake COVID-19 vaccination (hesitancy group). Differences in sociodemographic and health-related factors were compared with the COVID-19 vaccination hesitancy by using the chi-square statistics to determine the role of sociodemographic and health-related factors in COVID-19 vaccine hesitancy.

A multivariable analysis was developed in 2 stages. First, we performed a logistic regression to evaluate factors associated with COVID-19 vaccination hesitancy, including sociodemographic factors (ie, gender, age, family size, education, marital status, income, employment) and health-related factors (ie, subjective health and presence of underlying disease), media use (ie, offline media, online media, and social media), and psychological factors (ie, perceived susceptibility and severity toward COVID-19, perceived benefits and barriers of COVID-19 vaccination, and trust in government). In the second stage, a path analysis was carried out to describe the direct and indirect associations of media use and psychological factors with vaccine hesitancy using the Lavaan package (v0.6-9) [41] based on R version 4.0.5. The 3 types of media use (offline, online, and social media) were included in the path model at the same time to test the independent relations. Sociodemographic factors and health-related factors were added in the path model as control variables. Path models are a statistical method that, compared to multiple regressions, allow for the simultaneous assessment of several regression paths occurring between multiple dependent and independent variables and for the computing of direct, indirect (mediated), and total effects. Standardized parameter estimates were used to compare the magnitude of associations of the media use on mediators. Statistical analyses were conducted using R version 4.2 (R Foundation for Statistical Computing).

Results

Characteristics of the Survey Participants

Among the 1016 participants, 48.8% (496/1016) were men and 51.2% (520/1016) were women, with a mean age of 47.04 (SD 15.04) years (Table 1). The majority of participants had a family size of more than 2 persons (870/1016, 85.6%), and 59.9% (609/1016) were married. Half of the participants had at least some college education (532/1016, 52.4%), followed by those with only a high school education (456/1016, 44.9%). The most common monthly household income was approximately 2-3.99 million won (US $1688-US $3369; 360/1016, 35.4%), followed by 4-5.99 million won (US $3377-US $5057; 220/1016, 21.7%), and over 6 million won (US $5065; 52/1016, 2.4%) (Table 1). Among the participants, 87.3% (887/1016) lived in the urban areas, and about 22.2% (226/1016) had school-aged children. Additionally, 67% (681/1016) had received a seasonal influenza vaccination more than once in the previous 5 years, and 36.2% (368/1016) had more than one underlying disease. Approximately 35.5% (361/1016) reported their subjective health as good, 49.8% (506/1016) reported moderate, and 14.7% reported poor. Only 1.8% (18/1016) reported they had previously experienced COVID-19. Table 1 presents the characteristics of the sample population.

Table 1. Sociodemographic and health-related characteristics of the study participants (N=1016).

Characteristics	Values, n (%)
Sociodemographic factors	
Gender	
Male	496 (48.8)
Female	520 (51.2)
Age (years), mean 47.04 (SD 15.04) years	
18-29	170 (16.7)
30-39	157 (15.5)
40-49	190 (18.7)
50-59	201 (19.8)
≥60	298 (29.3)
Family size	
1 (living alone)	146 (14.4)
more than 2	870 (85.6)
Marital status	
Married	609 (59.9)
Single/divorced/bereaved	407 (4.1)
Presence of children	
None	790 (77.8)
More than 1	226 (22.2)
Education level	
Middle school or below	28 (2.8)
High school graduate	456 (44.9)
College	468 (46.1)
Graduate school and above	64 (6.3)
Income level (million won)[a]	
<2	229 (22.5)
2-3.99	360 (35.4)
4-5.99	220 (21.7)
6-7.99	155 (15.3)
≥8	52 (5.1)
Residence	
Urban	887 (87.3)
Rural	129 (12.7)
Health-related factors	
Seasonal influenza vaccination history (5 years)	
No	335 (32.9)
Yes	681 (67.0)
Underlying disease	
None	648 (63.8)
More than 1	368 (36.2)
Subjective health	
Poor	149 (14.7)

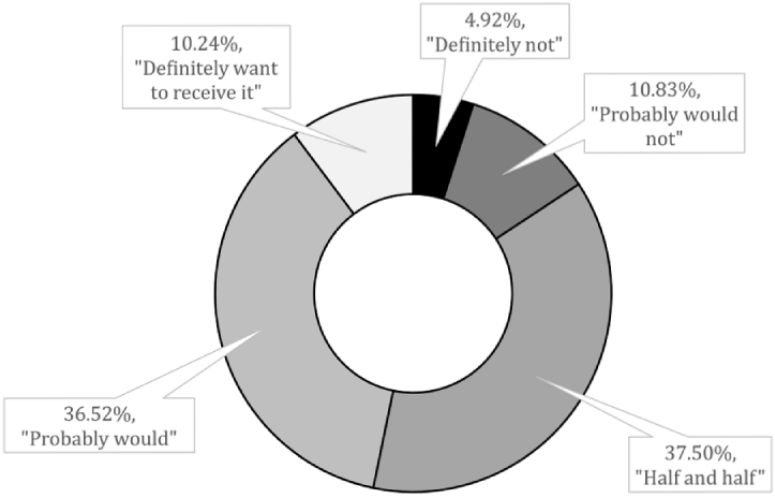
Characteristics	Values, n (%)
Moderate	506 (49.8)
Good	361 (35.5)
COVID-19 experience	
None	998 (98.2)
Confirmed	18 (1.8)

[a]Currency exchange conversion rate of 1000 won=US $0.87 is applicable.

Vaccine Hesitancy

Among the 1016 participants, 10.2% (104/1016) stated that they definitely would accept the vaccination, while 36.5% (371/1016) would probably accept the vaccination. However, 37.5% (381/1016) of the participants reported that they were "half and half," 10.8% (110/1016) would "probably would not," and 4.9% (50/1016) would "definitely not" get vaccinated (Figure 2). The COVID-19 vaccination intention response options of "definitely not," "probably would not," and "half and half" were grouped as "vaccine hesitancy" group (541/1016, 53.3%). The options "probably would" and "definitely want to receive it" were grouped as "vaccine acceptance" group (475/1016, 46.8%).

Figure 2. Distribution of responses regarding the question "If a vaccine for coronavirus (COVID-19) becomes available, would you want to receive it?"



Differences in the sociodemographic and health-related factors were compared with the COVID-19 vaccination hesitancy by using chi-square statistics (Table 2). Women (P<.001) and unmarried participants (single, divorced, or bereaved) (P<.001) were more likely to demonstrate vaccine hesitancy. Group differences by age (P<.001) and monthly household income (P=.03) were statistically significant. Health-related factors such as having an influenza vaccination history (P<.001) and the presence of underlying disease (P=.003) were related to higher vaccine acceptance. However, group differences between participants' different education levels and the presence of children were not statistically significant.

Table 2. Chi-square statistics for variables related to COVID-19 vaccination hesitancy and acceptance.

Characteristics	Likelihood of getting the COVID-19 vaccine (N=1016)		*P* value
	Acceptance (n=475), n (%)	Hesitancy (n=541), n (%)	
Sociodemographic factors			
Gender			<.001
Male	272 (54.8)	224 (45.2)	
Female	203 (39)	317 (61)	
Age (years)			<.001
18-29	55 (32.4)	115 (67.6)	
30-39	51 (32.5)	106 (67.5)	
40-49	91 (47.9)	99 (52.1)	
50-59	107 (53.2)	94 (46.8)	
≥60	171 (57.4)	127 (42.6)	
Family size			.89
1 (living alone)	69 (47.3)	77 (52.7)	
More than 2	406 (46.7)	464 (53.3)	
Education level			.97
Middle school or below	12 (42.9)	16 (57.1)	
High school graduate	214 (46.9)	242 (53.1)	
College	218 (46.6)	250 (53.4)	
Graduate school and above	31 (48.4)	33 (51.6)	
Marital status			<.001
Married	312 (51.2)	297 (48.8)	
Single/divorced/bereaved	163 (40)	244 (60)	
Presence of children			.39
None	375 (47.5)	415 (52.5)	
More than 1	100 (44.2)	126 (55.8)	
Income level (million won)[a]			.03
<2	103 (45)	126 (55)	
2-3.99	164 (45.6)	196 (54.4)	
4-5.99	113 (51.4)	107 (48.6)	
6-7.99	80 (51.6)	75 (48.4)	
≥8	15 (28.8)	37 (71.2)	
Residence			.61
Urban	412 (46.4)	475 (53.6)	
Rural	63 (48.8)	66 (51.2)	
Health-related factors			
Influenza vaccination history			<.001
No	107 (31.9)	228 (68.1)	
Yes	368 (54)	313 (46)	
Underlying disease			.003
None	280 (43.2)	368 (56.8)	
More than 1	195 (53)	173 (47)	
Subjective health			.14

Characteristics	Likelihood of getting the COVID-19 vaccine (N=1016)		*P* value
	Acceptance (n=475), n (%)	Hesitancy (n=541), n (%)	
Poor	79 (53)	70 (47)	
Moderate	223 (44.1)	283 (55.9)	
Good	173 (47.9)	188 (52.1)	
COVID-19 infection experience			.50
Not infected	468 (46.9)	530 (53.1)	
Confirmed	7 (38.9)	11 (61.1)	

[a]Currency exchange conversion rate of 1000 won=US $0.87 is applicable.

Media Used to Obtain COVID-19 Vaccination Information

We examined how often people used media sources (offline media, online media, social media) to learn about COVID-19 vaccination. Interestingly, participants sought little vaccine-related health information via social media; the average social media information seeking was close to "sometimes" (score=2) (mean 1.83 [SD 0.70]). By contrast, the participants used online media more often (mean 2.70 [SD 0.81]), followed by offline media (mean 2.66 [SD 0.95]) (Table 3).

Table 3. Media use and psychological characteristics of the study participants.

Characteristics	Values, mean (SD)
Media use (4-point scale)	
Offline media	2.66 (0.95)
Online media	2.70 (0.81)
Social media	1.83 (0.70)
Psychological factors (5-point scale)	
Perceived susceptibility	3.05 (0.74)
Perceived severity	3.95 (0.76)
Perceived benefits	3.38 (0.81)
Self-protection of my health	3.74 (0.97)
Proved efficacy of the vaccine	2.80 (0.96)
Perceived barriers	3.29 (0.90)
Vaccination caused COVID-19 infection	2.85 (1.11)
Concerns about vaccination's side effects	3.72 (1.01)
Trust in government	2.80 (0.68)

Psychological Factors Related to COVID-19 Vaccination

Participants perceived the risk of becoming infected with COVID-19 (perceived susceptibility) as being "moderate" (score=3) (mean 3.05 [SD 0.74]). Only 3.1% (31/1016) reported that perceived chance of infection is "very high" (score=5) and 18.8% (191/1016) reported "high" (score=4). Many participants reported that the chance of infection is "neither high nor low" (621/1016, 61.1%). The average perceived severity score was higher than perceived susceptibility, which was close to "high" (score=4) (mean 3.95 [SD 0.76]). However, among the participants, 55.1% (560/1016) reported that the severity would be "high" (score=4), and 22% (195/1016) reported "very high" (score=5). Participants' perception of vaccination benefits was measured by 2 items measuring perceived chances of gaining specific benefits by COVID-19 vaccination. The average score was higher than "moderate" (score=3) (mean 3.38 [SD 0.81]). Among the benefits, "self-protection of my health" was the highest (mean 3.74 [SD 0.968]) and "proved efficacy of the vaccine" was the lowest (mean 2.80 [SD 0.964]). The average score for perceived barriers to vaccination was neither high nor low (mean 3.29 [SD 0.90]). The belief that a vaccination caused COVID-19 infection was relatively low (mean 2.85 [SD 1.110]); however, concerns about side effects were high (mean 3.72 [SD 1.007]). Among the participants, 23.8% (242/1016) reported their concerns that the vaccination's side effects are "very high," while 38.2% (388/1016) reported they were "high." The average score of trust in government was slightly less than moderate (mean 2.80 [SD 0.68]) (Table 3).

Factors Associated With COVID-19 Vaccine Hesitancy

Table 4 shows the results of the hierarchical logit regression models to test the association between vaccine hesitancy and

participants' sociodemographic factors, health-related factors, media use, and psychological factors related to COVID-19 vaccine hesitancy. A total of 60.5% of the variance was explained by the final model. The Nagelkerke R^2 changes indicated that the incremental variances explained by each block of variables were 16.7%, 3.10%, and 40.7% for sociodemographic and health-related characteristics, media use, psychological responses, respectively. Out of the sociodemographic factors, female (odds ratio [OR] 1.967, 95% CI 1.36-2.86; *P*<.001), age in 40s (OR 0.467, 95% CI 0.23-0.95; *P*=.003), 50s (OR 0.47, 95% CI 0.23-0.96; *P*=.004), and over 60s (OR 0.49, 95% CI 0.24-0.99; *P*=.04) were significant individual predictors of COVID-19 vaccine hesitancy. Among health-related factors, participants who had seasonal influenza vaccination in 5 years (OR 0.50, 95% CI 0.34-0.75; *P*<.001) were less likely to hesitate to uptake the COVID-19 vaccine. Seeking COVID-19 vaccine-related information via social medias was related to higher tendency of vaccine hesitancy (OR 1.46, 95% CI 1.10-1.92; *P*=.01). Regarding psychological factors related to COVID-19 vaccination, among perceived risks of COVID-19 infection, higher perceived susceptibility was associated with lower vaccine hesitancy (OR 0.69, 95% CI 0.52-0.91; *P*=.01). Regarding the perceived benefits and barriers of COVID-19 vaccination, perceived benefits were related to lower chance of vaccine hesitancy (OR 0.007, 95% CI 0.05-0.10; *P*<.001) while perceived barriers (OR 1.63, 95% CI 1.29-2.07; *P*<.001) were related to vaccine hesitancy. Finally, lower trust in government was associated with vaccine hesitancy significantly (OR 0.72, 95% CI 0.53-0.98; *P*=.04).

Table 4. Factors associated with COVID-19 vaccination hesitancy[a].

Characteristics	Odds ratio (95% CI)	*P* value
Gender		
Male	Ref[b]	
Female	1.967 (1.353-2.86)	<.001
Age (years)		
18-29	Ref	
30-39	1.052 (0.538-2.059)	.88
40-49	0.467 (0.23-0.945)	.03
50-59	0.47 (0.229-0.964)	.04
≥60	0.49 (0.239-0.99)	.04
Education level		
Under middle school	Ref	
High school graduate	0.373 (0.12-1.162)	.09
College	0.293 (0.092-0.929)	.04
Graduate school	0.377 (0.099-1.438)	.15
Income level (million won)[c]		
<2	Ref	
2-3.99	1.114 (0.681-1.824)	.67
4-5.99	0.871 (0.489-1.553)	.64
6-7.99	0.923 (0.485-1.758)	.81
≥8	2.09 (0.816-5.353)	.12
Marital status		
Single/divorced/bereaved	Ref	
Married	1.099 (0.642-1.882)	.73
Presence of children		
None	Ref	
More than 1	1.504 (0.853-2.652)	.16
Residential area		
Urban	Ref	
Town	1.324 (0.752-2.331)	.33
Influenza vaccination history		
No	Ref	
Yes	0.501 (0.337-0.746)	.001
Underlying disease		
None	Ref	
More than 1	1.213 (0.796-1.846)	.37
Subjective health		
Bad	Ref	
Moderate	0.893 (0.50-1.596)	.70
Good	0.999 (0.533-1.872)	.99
COVID-19 infection experience		
Not infected	Ref	

Characteristics	Odds ratio (95% CI)	*P* value
Confirmed	3.419 (0.836-13.982)	.09
Media use		
Social media	1.455 (1.101-1.922)	.008
Online media	0.838 (0.616-1.139)	.26
Offline media	0.939 (0.736-1.197)	.61
Psychological factors		
Perceived susceptibility	0.685 (0.516-0.909)	.009
Perceived severity	0.898 (0.679-1.187)	.45
Perceived benefits	0.067 (0.045-0.099)	<.001
Perceived barriers	1.631 (1.285-2.069)	<.001
Trust in government	0.719 (0.528-0.978)	.04

[a]Nagelkerke R^2=0.61.

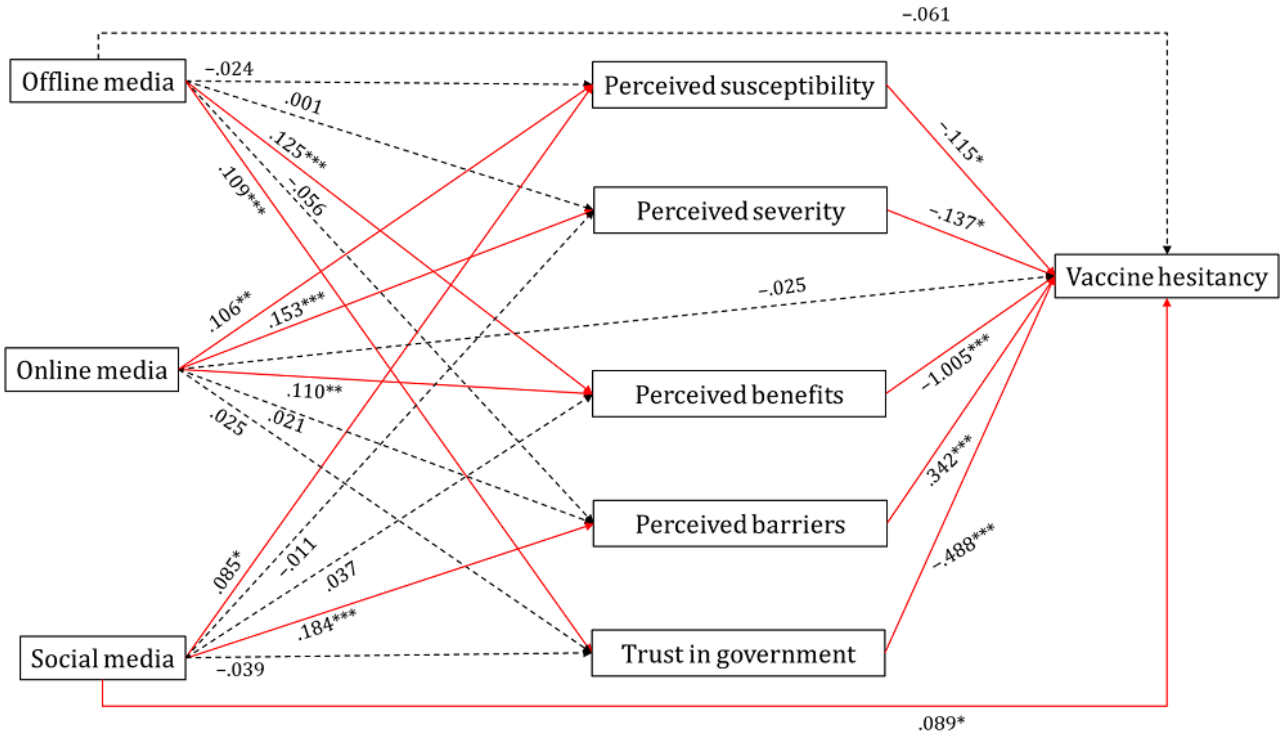
[b]Ref: reference value.

[c]Currency exchange conversion rate of 1000 won=US $0.87 is applicable.

Indirect Association Between Media Use and Vaccine Hesitancy Via Psychological Factors

The results of the path model analysis showing the relationships between media use, psychological factors, and vaccine hesitancy are shown in Figure 3. Higher level of perceived susceptibility of COVID-19 was related with using online media (β=.106, *P*=.004) and social media (β=.085, *P*=.01) for vaccine-related information, while perceived severity of the disease was related to online media use (β=.153, *P*<.001). Regarding perceived benefits of COVID-19 vaccination, offline media use (β=.125, *P*<.001) and online media use (β=.110, *P*=.002) were related in a positive direction. Higher perception on barriers of vaccination was associated with social media use (β=.184, *P*<.001). Lastly, offline media use (β=.109, *P*<.001) was related to higher trust in government (Table 5).

Figure 3. Path model showing the relationships between media use, psychological factors, and vaccine hesitancy. Note: Path coefficients are standardized regression weights. Nonsignificant paths are indicated as dotted lines. *P<.005, **P<.01, ***P<.001.



We examined the standardized parameter estimates to compare the magnitude of associations of media use on vaccine hesitancy via psychological factors (ie, perceived susceptibility and severity of COVID-19 infection, perceived benefits and barriers of COVID-19 vaccination, and trust in government). Offline media use was related to lower vaccine hesitancy indirectly via higher perceived benefits and trust in government. With regard to online media use, the indirect association through perceived

severity of COVID-19 infection and perceived benefits of vaccines were significant and related to lower vaccine hesitancy. Lastly, the indirect association between social media and vaccine hesitancy via perceived susceptibility (P=.01) of COVID-19 and perceived barriers of vaccines (P<.001) was significant. Interestingly, we found conflicting indirect relations between social media use and vaccine hesitancy. Although perceived susceptibility of COVID-19 infection negatively mediated the relationship of social media use with vaccine hesitancy, perceived barriers mediated the relationship of social media use positively.

Table 5. Standardized parameter estimates of indirect association between media use and vaccine hesitancy via mediators.

Independent variable, mediator	Vaccine hesitancy		
	Standardized parameter estimates	SE	*P* value
Offline media			
Perceived susceptibility	0.003	0.004	.46
Perceived severity	0.001	0.004	.99
Perceived benefits	–0.126	0.031	<.001
Perceived barriers	–0.019	0.013	.14
Trust in government	–0.053	0.015	<.001
Online media			
Perceived susceptibility	–0.012	0.007	.09
Perceived severity	–0.021	0.01	.04
Perceived benefits	–0.111	0.036	.002
Perceived barriers	0.007	0.015	.62
Trust in government	–0.012	0.017	.49
Social media			
Perceived susceptibility	–0.01	0.006	.01
Perceived severity	0.002	0.005	.77
Perceived benefits	–0.037	0.037	.31
Perceived barriers	0.063	0.016	<.001
Trust in government	0.019	0.015	.22

Discussion

Principal Results

This study revealed a considerable level of COVID-19 vaccine hesitancy at 1 month before the start of vaccination in South Korea. Among the participants, 53.3% (541/1016) hesitated to uptake the COVID-19 vaccination, 37.5% (381/1016) reported they were “half and half” and 10.8% (110/1016) reported they would “probably not” or “definitely not” get vaccinated, while 46.7% (475/1016) of the participants would “likely” or “very likely” accept the vaccine. A decrement of vaccine acceptance was also observed in South Korea, as in other countries, from 79.8% (June 2020) to 46.7% (January 2021) [4-6]. This shows that efforts to increase vaccine acceptance are needed to meet the South Korean government’s target of 70% of the total population. Interventions and policies to raise vaccine acceptance and uptake are urgently needed.

Several additional findings are worthy of note. First, sociodemographic characteristics and health-related factors were standard subgroup variables cross-tabulated with vaccination hesitancy. Gender has emerged as a significant issue during this pandemic. Previous studies have shown that women are more likely to engage in preventive behaviors [42-44]; however, they are less willing than men to receive the vaccine [29], with more females declaring that they are unsure of taking the vaccine. Similarly, our study results reveal that women (224/1016, 45.2%) are more hesitant to receive COVID-19 vaccination than men (317/1016, 61%). Regarding age, higher levels of hesitancy among the participants in their 20s (115/170, 67.6%) and 30s (106/157, 67.5%) leave them particularly vulnerable to COVID-19. This result is similar to prior research investigating the relationship between sociodemographic factors and vaccine hesitancy during the COVID-19 pandemic [4,45,46]. Thus, gender-based and generation-based public health policies and communication are recommended.

Second, the results revealed the association between psychological factors and vaccine hesitancy. Perceived barriers such as concerns about side effects caused by a COVID-19 vaccine had a significant and robust association with vaccine hesitancy. The perceived susceptibility of COVID-19 infection and benefits of vaccination (eg, proven efficacy of COVID-19 vaccines) related negatively to vaccine hesitancy. Trust in government was also negatively associated with vaccine hesitancy. The predictive power of psychological factors on vaccine hesitancy has been emphasized by numerous studies

[21,47,48]; similarly, psychological factors explained 40.7% of the variance for vaccine hesitancy in this study. Therefore, interventions and communication strategies aimed to improve perceived benefits of vaccination and reducing perceived barriers to vaccination will be useful in responding to vaccine hesitancy [48]. A well-known aspect of COVID-19 vaccination is that it may induce adverse events. According to a report of COVID-19 vaccine safety monitoring in South Korea, 16,196 adverse events (0.5%) following 3,586,814 administered doses of COVID-19 vaccines were reported in approximately 2 months (February 26 to April 30, 2021) [49]. Of these, 15,658 (96.7%) were nonserious adverse events comprising local and systemic reactions, including myalgia, headache, fever, and pain at the injection site; however, 538 (3.3%) were serious adverse events, including 73 (0.5%) deaths. These results are similar to the findings of the clinical trial [50]. Although possible side effects are comparatively rare, the risk of their occurrence may significantly influence the choice to not vaccinate. The decision to vaccinate (or not) shares many common points with one's chosen level of self-protection [51]. The most obvious link between the two problems is that the decision maker's objective is to reduce the probability of an undesirable event. Vaccination reduces the probability of the primary disease; however, another cost may induce other new risks. From this point of view, the decision-making on vaccination is a trade-off between the risk of the primary disease and the risk of incurring side effects [52]. Therefore, the vaccine hesitancy due to the perceived barrier such as concerns on vaccine-induced side effects can be regarded as a risk aversion [53,54]. Moreover, in Korea, face masks were promoted based on substantial relative benefits, high efficacy of slowing viral spread, and low cost. Several studies report high compliance and high efficacy beliefs on wearing masks among the Korean population [42-44]. Therefore, people might prefer to wear face masks over vaccination to avoid the risk of the side effects from vaccination.

Third, this study suggests that the use of media to obtain vaccine-related information has a positive or negative relation with the population's vaccination decision-making. The mechanisms by which psychological factors mediate the association between media use and vaccine hesitancy were also revealed. Offline media such as TV, radio, and newspapers were associated with higher perceived benefits of a COVID-19 vaccine and higher trust in government, which led to lower vaccine hesitancy. Kim and Jung [28] suggested that appropriate use of offline media (eg, radio and newspapers) is critical to increase the population's vaccination rate and is a way to reduce the information gap among social classes. Online media also had an indirect association with vaccine hesitancy in a negative direction, mediated through perceived severity of COVID-19 infection and perceived benefits of the vaccine. This result is contradictory with that reported in other studies reporting the association between online media use and higher vaccine hesitancy [31-33]. Therefore, further studies to examine the role of online media on vaccine hesitancy are needed.

Regarding social media, frequent social media use was related to vaccine hesitancy even when other types of media use (eg, offline, online media) and sociodemographic, health-related, and psychological factors were included in the model. Social media use was related with vaccine hesitancy via perceived susceptibility in a negative direction and perceived barriers in a positive direction. This result implies that social media as a health information source could act as a double-edged sword. Social media has been criticized because it propagates more misinformation than any other media type [55] and induces a high level of anxiety, depression, and even conspiracy beliefs during the COVID-19 pandemic [56,57]. The vaccine hesitancy owing to false beliefs caused by controversies and misinformation in social media is well documented in the literature [30,39,58,59]. Nevertheless, social media can also be used to build a positive perception of vaccination [30] and disseminate valuable evidence-based health information and recommendations rapidly to many people timely [60]. The quality and reliability of the information provided via social media should be improved [34,61]; public health communicators are encouraged to establish a web-based reputation as experts worth following or visiting online [30].

Implications

A set of implications for interventions, communication strategies, and future research can be drawn from this study. Since this study was conducted a month before the start of vaccination, it provides a valuable opportunity to understand the public's responses on novel vaccines and develop practical implications for implementing effective vaccination programs in a future epidemic. First, as the perceived barrier was the highest predictor of vaccine hesitancy, efforts to lower the perceived risk of vaccine side effects and heighten perceived benefits of the vaccine are required. This study indicates that social media is related to heightened perceived barriers, thereby leading to vaccine hesitancy, which is not very surprising. The phenomenon of "infodemics," defined as the rapid spread and amplification of vast amounts of valid and invalid information on the internet or through other media, is a tremendous and ongoing challenge in the COVID-19 pandemic [62]. Communication efforts of public health authorities to provide accurate and reliable information, confront misinformation or disinformation, and reduce the negative impact of such infodemics are required. Eysenbach [63] recommended in his commentary to promote (1) information monitoring, (2) health and science literacy skills, (3) fact checking or peer review of the information, and (4) timely and accurate knowledge transition to fight the infodemics. Using alternative communication tools can provide official messages from the public health authorities to the public. For instance, emergency alert text messages [44] or official social media accounts [37] could be considered. Second, building trust in the government and public health authorities offers another solution. Trust is an essential factor influencing people's perception and behavior during a pandemic and influences people's willingness to be vaccinated [64,65]. Trust in institutions or persons decreases the perceived risk of new technologies and indirectly impacts the higher risk acceptance of the new technologies [64,66]. As the COVID-19 vaccine is newly developed and the development period was short, the public might recognize the COVID-19 vaccine as new technology. Therefore, it could be assumed that the public trust in governments can encourage the public to accept the risk of adverse events induced by COVID-19

vaccination. Trust is critical, and it is more easily broken and difficult to maintain under high-risk and high-uncertainty situations [67].

Limitations

This study has several limitations worthy of note. First, although there are several types of social media such as official social media, professional social media, and public social media [37], this study did not examine the effect of each type of social media. Further studies should examine the effect of each type of social media and whether the effect is mediated via psychological factors. Second, we were not able to investigate the potential psychological factors proposed by previous studies such as subjective norms [68,69], knowledge [70], or contextual barriers such as accessibility [71]. Finally, since this study was designed cross-sectionally, we identified the associations between media use, psychological factors, and vaccine hesitancy rather than causal inference.

Conclusions

This study's findings suggest that social media plays a role in disseminating vaccine-related information, especially vaccination benefits, which can help the public accept the COVID-19 vaccine for disease control. Conversely, social media can also increase concerns about a vaccination's side effects, stimulating vaccine hesitancy. Sociodemographic factors such as gender and age should be considered in public health interventions, and greater attention should be given to the younger adults and females during a pandemic for effective vaccination campaigns. This study highlights the need for government and public health authorities to consider approaching their public with additional efforts to promote acceptance of the COVID-19 vaccine. Efforts to disseminate reliable and timely information and monitor social media misinformation are needed during the pandemic. Understanding the intentions of vaccination and factors associated with the intentions may help inform public health authorities about evidence-based interventions; vaccination strategies are necessary to achieve broader community uptake.

Acknowledgments

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2021S1A5B5A16075887).

Authors' Contributions

ML and MY conceptualized this study. ML designed the methodology for this study, conducted formal analysis, and was involved in the writing and initial draft preparation. MY acquired the data.

Conflicts of Interest

None declared.

References

1. Nelson C, Lurie N, Wasserman J, Zakowski S. Conceptualizing and Defining Public Health Emergency Preparedness. Am J Public Health 2007 Apr;97(Supplement_1):S9-S11. [doi: 10.2105/ajph.2007.114496]
2. Anderson RM, May RM. Vaccination and herd immunity to infectious diseases. Nature 1985;318(6044):323-329. [doi: 10.1038/318323a0] [Medline: 3906406]
3. Jung J. Preparing for the Coronavirus Disease (COVID-19) Vaccination: Evidence, Plans, and Implications. J Korean Med Sci 2021 Feb 22;36(7):e59 [FREE Full text] [doi: 10.3346/jkms.2021.36.e59] [Medline: 33619920]
4. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
5. Sallam M. COVID-19 Vaccine Hesitancy Worldwide: A Concise Systematic Review of Vaccine Acceptance Rates. Vaccines (Basel) 2021 Feb 16;9(2):160 [FREE Full text] [doi: 10.3390/vaccines9020160] [Medline: 33669441]
6. Lin C, Tu P, Beitsch LM. Confidence and Receptivity for COVID-19 Vaccines: A Rapid Systematic Review. Vaccines (Basel) 2020 Dec 30;9(1):16 [FREE Full text] [doi: 10.3390/vaccines9010016] [Medline: 33396832]
7. Benin AL, Wisler-Scher DJ, Colson E, Shapiro ED, Holmboe ES. Qualitative analysis of mothers' decision-making about vaccines for infants: the importance of trust. Pediatrics 2006 May;117(5):1532-1541. [doi: 10.1542/peds.2005-1728] [Medline: 16651306]
8. Keane MT, Walter MV, Patel BI, Moorthy S, Stevens RB, Bradley KM, et al. Confidence in vaccination: a parent model. Vaccine 2005 Mar 31;23(19):2486-2493. [doi: 10.1016/j.vaccine.2004.10.026] [Medline: 15752835]
9. Nichter M. Vaccinations in the third world: A consideration of community demand. Social Science & Medicine 1995 Sep;41(5):617-632. [doi: 10.1016/0277-9536(95)00034-5]
10. MacDonald NE, SAGE Working Group on Vaccine Hesitancy. Vaccine hesitancy: Definition, scope and determinants. Vaccine 2015 Aug 14;33(34):4161-4164 [FREE Full text] [doi: 10.1016/j.vaccine.2015.04.036] [Medline: 25896383]

11. Dubé E, Laberge C, Guay M, Bramadat P, Roy R, Bettinger JA. Vaccine hesitancy: an overview. Hum Vaccin Immunother 2013 Aug;9(8):1763-1773 [FREE Full text] [doi: 10.4161/hv.24657] [Medline: 23584253]
12. Opel DJ, Mangione-Smith R, Taylor JA, Korfiatis C, Wiese C, Catz S, et al. Development of a survey to identify vaccine-hesitant parents: the parent attitudes about childhood vaccines survey. Hum Vaccin 2011 Apr;7(4):419-425 [FREE Full text] [doi: 10.4161/hv.7.4.14120] [Medline: 21389777]
13. Bulut C, Kato Y. Epidemiology of COVID-19. Turk J Med Sci 2020 Apr 21;50(SI-1):563-570 [FREE Full text] [doi: 10.3906/sag-2004-172] [Medline: 32299206]
14. Torequl Islam M, Nasiruddin M, Khan IN, Mishra SK, Kudrat-E-Zahan M, Alam Riaz T, et al. A Perspective on Emerging Therapeutic Interventions for COVID-19. Front Public Health 2020;8:281 [FREE Full text] [doi: 10.3389/fpubh.2020.00281] [Medline: 32733837]
15. El-Shitany NA, Harakeh S, Badr-Eldin SM, Bagher AM, Eid BG, Almukadi HS, et al. Minor to Moderate Side Effects of Pfizer-BioNTech COVID-19 Vaccine Among Saudi Residents: A Retrospective Cross-Sectional Study. IJGM 2021 Apr;Volume 14:1389-1401. [doi: 10.2147/ijgm.s310497]
16. Kadali RA, Janagama R, Peruru S, Malayala SV. Side effects of BNT162b2 mRNA COVID-19 vaccine: A randomized, cross-sectional study with detailed self-reported symptoms from healthcare workers. Int J Infect Dis 2021 May;106:376-381 [FREE Full text] [doi: 10.1016/j.ijid.2021.04.047] [Medline: 33866000]
17. Østergaard SD, Schmidt M, Horváth-Puhó E, Thomsen RW, Sørensen HT. Thromboembolism and the Oxford–AstraZeneca COVID-19 vaccine: side-effect or coincidence? The Lancet 2021 Apr;397(10283):1441-1443. [doi: 10.1016/s0140-6736(21)00762-5]
18. Caserotti M, Girardi P, Rubaltelli E, Tasso A, Lotto L, Gavaruzzi T. Associations of COVID-19 risk perception with vaccine hesitancy over time for Italian residents. Soc Sci Med 2021 Mar;272:113688 [FREE Full text] [doi: 10.1016/j.socscimed.2021.113688] [Medline: 33485215]
19. Sherman SM, Smith LE, Sim J, Amlôt R, Cutts M, Dasch H, et al. COVID-19 vaccination intention in the UK: results from the COVID-19 vaccination acceptability study (CoVAccS), a nationally representative cross-sectional survey. Hum Vaccin Immunother 2021 Jun 03;17(6):1612-1621 [FREE Full text] [doi: 10.1080/21645515.2020.1846397] [Medline: 33242386]
20. Soares P, Rocha JV, Moniz M, Gama A, Laires PA, Pedro AR, et al. Factors Associated with COVID-19 Vaccine Hesitancy. Vaccines (Basel) 2021 Mar 22;9(3):300 [FREE Full text] [doi: 10.3390/vaccines9030300] [Medline: 33810131]
21. Murphy J, Vallières F, Bentall RP, Shevlin M, McBride O, Hartman TK, et al. Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. Nat Commun 2021 Jan 04;12(1):29 [FREE Full text] [doi: 10.1038/s41467-020-20226-9] [Medline: 33397962]
22. Schernhammer E, Weitzer J, Laubichler M, Birmann B, Bertau M, Zenk L, et al. Correlates of COVID-19 vaccine hesitancy in Austria: trust and the government. J Public Health (Oxf) 2021 May 05:2021 [FREE Full text] [doi: 10.1093/pubmed/fdab122] [Medline: 33948665]
23. Chen H, Li X, Gao J, Liu X, Mao Y, Wang R, et al. Health Belief Model Perspective on the Control of COVID-19 Vaccine Hesitancy and the Promotion of Vaccination in China: Web-Based Cross-sectional Study. J Med Internet Res 2021 Sep 06;23(9):e29329 [FREE Full text] [doi: 10.2196/29329] [Medline: 34280115]
24. Fisher KA, Bloomstone SJ, Walder J, Crawford S, Fouayzi H, Mazor KM. Attitudes Toward a Potential SARS-CoV-2 Vaccine. Ann Intern Med 2020 Dec 15;173(12):964-973. [doi: 10.7326/m20-3569]
25. Kreps S, Prasad S, Brownstein JS, Hswen Y, Garibaldi BT, Zhang B, et al. Factors Associated With US Adults' Likelihood of Accepting COVID-19 Vaccination. JAMA Netw Open 2020 Oct 01;3(10):e2025594 [FREE Full text] [doi: 10.1001/jamanetworkopen.2020.25594] [Medline: 33079199]
26. Machida M, Nakamura I, Kojima T, Saito R, Nakaya T, Hanibuchi T, et al. Acceptance of a COVID-19 Vaccine in Japan during the COVID-19 Pandemic. Vaccines (Basel) 2021 Mar 03;9(3):210 [FREE Full text] [doi: 10.3390/vaccines9030210] [Medline: 33802285]
27. Sallam M, Dababseh D, Eid H, Al-Mahzoum K, Al-Haidar A, Taim D, et al. High Rates of COVID-19 Vaccine Hesitancy and Its Association with Conspiracy Beliefs: A Study in Jordan and Kuwait among Other Arab Countries. Vaccines (Basel) 2021 Jan 12;9(1):42 [FREE Full text] [doi: 10.3390/vaccines9010042] [Medline: 33445581]
28. Kim J, Jung M. Associations between media use and health information-seeking behavior on vaccinations in South Korea. BMC Public Health 2017 Sep 11;17(1):700 [FREE Full text] [doi: 10.1186/s12889-017-4721-x] [Medline: 28893212]
29. Piltch-Loeb R, Savoia E, Goldberg B, Hughes B, Verhey T, Kayyem J, et al. Examining the effect of information channel on COVID-19 vaccine acceptance. PLoS One 2021;16(5):e0251095 [FREE Full text] [doi: 10.1371/journal.pone.0251095] [Medline: 33979370]
30. Hwang J. Health Information Sources and the Influenza Vaccination: The Mediating Roles of Perceived Vaccine Efficacy and Safety. J Health Commun 2020 Sep 01;25(9):727-735. [doi: 10.1080/10810730.2020.1840675] [Medline: 33186091]
31. Weiner JL, Fisher AM, Nowak GJ, Basket MM, Gellin BG. Childhood Immunizations: First-Time Expectant Mothers' Knowledge, Beliefs, Intentions, and Behaviors. Am J Prev Med 2015 Dec;49(6 Suppl 4):S426-S434 [FREE Full text] [doi: 10.1016/j.amepre.2015.07.002] [Medline: 26297448]
32. Witteman HO, Zikmund-Fisher BJ. The defining characteristics of Web 2.0 and their potential influence in the online vaccination debate. Vaccine 2012 May 28;30(25):3734-3740. [doi: 10.1016/j.vaccine.2011.12.039] [Medline: 22178516]

33. Park S, Massey PM, Stimpson JP. Primary Source of Information About COVID-19 as a Determinant of Perception of COVID-19 Severity and Vaccine Uptake : Source of Information and COVID-19. J Gen Intern Med 2021 Oct;36(10):3088-3095 [FREE Full text] [doi: 10.1007/s11606-021-07080-1] [Medline: 34378115]
34. Benis A, Khodos A, Ran S, Levner E, Ashkenazi S. Social Media Engagement and Influenza Vaccination During the COVID-19 Pandemic: Cross-sectional Survey Study. J Med Internet Res 2021 Mar 16;23(3):e25977 [FREE Full text] [doi: 10.2196/25977] [Medline: 33651709]
35. Graham MW, Avery EJ, Park S. The role of social media in local government crisis communications. Public Relations Review 2015 Sep;41(3):386-394. [doi: 10.1016/j.pubrev.2015.02.001]
36. McKee M, Cole K, Hurst L, Aldridge RW, Horton R. The other Twitter revolution: how social media are helping to monitor the NHS reforms. BMJ 2011 Feb 16;342:d948. [doi: 10.1136/bmj.d948] [Medline: 21325389]
37. Li X, Liu Q. Social Media Use, eHealth Literacy, Disease Knowledge, and Preventive Behaviors in the COVID-19 Pandemic: Cross-Sectional Study on Chinese Netizens. J Med Internet Res 2020 Oct 09;22(10):e19684 [FREE Full text] [doi: 10.2196/19684] [Medline: 33006940]
38. Puri N, Coomes EA, Haghbayan H, Gunaratne K. Social media and vaccine hesitancy: new updates for the era of COVID-19 and globalized infectious diseases. Hum Vaccin Immunother 2020 Nov 01;16(11):2586-2593 [FREE Full text] [doi: 10.1080/21645515.2020.1780846] [Medline: 32693678]
39. Wilson SL, Wiysonge C. Social media and vaccine hesitancy. BMJ Glob Health 2020 Oct 23;5(10):e004206 [FREE Full text] [doi: 10.1136/bmjgh-2020-004206] [Medline: 33097547]
40. Romer D, Jamieson KH. Patterns of Media Use, Strength of Belief in COVID-19 Conspiracy Theories, and the Prevention of COVID-19 From March to July 2020 in the United States: Survey Study. J Med Internet Res 2021 Apr 27;23(4):e25215 [FREE Full text] [doi: 10.2196/25215] [Medline: 33857008]
41. Rosseel Y. Lavaan: An R Package for Structural Equation Modeling. J Stat Soft 2012;48(2):A. [doi: 10.18637/jss.v048.i02]
42. Lee M, Kang B, You M. Knowledge, attitudes, and practices (KAP) toward COVID-19: a cross-sectional study in South Korea. BMC Public Health 2021 Feb 05;21(1):295 [FREE Full text] [doi: 10.1186/s12889-021-10285-y] [Medline: 33546644]
43. Lee M, You M. Psychological and Behavioral Responses in South Korea During the Early Stages of Coronavirus Disease 2019 (COVID-19). Int J Environ Res Public Health 2020 Apr 25;17(9):2977 [FREE Full text] [doi: 10.3390/ijerph17092977] [Medline: 32344809]
44. Lee M, You M. Effects of COVID-19 Emergency Alert Text Messages on Practicing Preventive Behaviors: Cross-sectional Web-Based Survey in South Korea. J Med Internet Res 2021 Feb 25;23(2):e24165. [doi: 10.2196/24165]
45. Neumann-Böhme S, Varghese NE, Sabat I, Barros PP, Brouwer W, van Exel J, et al. Once we have it, will we use it? A European survey on willingness to be vaccinated against COVID-19. Eur J Health Econ 2020 Sep 26;21(7):977-982 [FREE Full text] [doi: 10.1007/s10198-020-01208-6] [Medline: 32591957]
46. Wood S, Schulman K. Beyond Politics — Promoting Covid-19 Vaccination in the United States. N Engl J Med 2021 Feb 18;384(7):e23. [doi: 10.1056/nejmms2033790]
47. Cordina M, Lauri MA, Lauri J. Attitudes towards COVID-19 vaccination, vaccine hesitancy and intention to take the vaccine. Pharm Pract (Granada) 2021 Mar 21;19(1):2317. [doi: 10.18549/pharmpract.2021.1.2317]
48. Schmid P, Rauber D, Betsch C, Lidolt G, Denker M. Barriers of Influenza Vaccination Intention and Behavior - A Systematic Review of Influenza Vaccine Hesitancy, 2005 - 2016. PLoS One 2017;12(1):e0170550 [FREE Full text] [doi: 10.1371/journal.pone.0170550] [Medline: 28125629]
49. Oh H, Kim EK, Hwang I, Kim TE, Lee Y, Lee E, et al. COVID-19 vaccine safety monitoring in the Republic of Korea: February 26, 2021 to April 30, 2021. Osong Public Health Res Perspect 2021 Aug;12(4):264-268 [FREE Full text] [doi: 10.24171/j.phrp.2021.0157] [Medline: 34465075]
50. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, Türeci, et al. Safety and Efficacy of the BNT162b2 mRNA Covid-19 Vaccine. N Engl J Med 2020 Dec 31;383(27):2603-2615. [doi: 10.1056/nejmoa2034577]
51. Ehrlich I, Becker GS. Market Insurance, Self-Insurance, and Self-Protection. Journal of Political Economy 1972 Jul;80(4):623-648. [doi: 10.1086/259916]
52. Crainich D, Eeckhoudt L, Menegatti M. Vaccination as a trade-off between risks. Ital Econ J 2019 Mar 23;5(3):455-472. [doi: 10.1007/s40797-019-00089-w]
53. Halek M, Eisenhauer JG. Demography of Risk Aversion. The Journal of Risk and Insurance 2001 Mar;68(1):1. [doi: 10.2307/2678130]
54. Holt CA, Laury SK. Risk Aversion and Incentive Effects. American Economic Review 2002 Nov 01;92(5):1644-1655. [doi: 10.1257/000282802762024700]
55. Kouzy R, Abi Jaoude J, Kraitem A, El Alam MB, Karam B, Adib E, et al. Coronavirus Goes Viral: Quantifying the COVID-19 Misinformation Epidemic on Twitter. Cureus 2020 Mar 13;12(3):e7255 [FREE Full text] [doi: 10.7759/cureus.7255] [Medline: 32292669]
56. Allington D, Duffy B, Wessely S, Dhavan N, Rubin J. Health-protective behaviour, social media usage and conspiracy belief during the COVID-19 public health emergency. Psychol. Med 2020 Jun 09;51(10):1763-1769. [doi: 10.1017/s003329172000224x]

57. Lee JJ, Kang K, Wang MP, Zhao SZ, Wong JYH, O'Connor S, et al. Associations Between COVID-19 Misinformation Exposure and Belief With COVID-19 Knowledge and Preventive Behaviors: Cross-Sectional Online Study. J Med Internet Res 2020 Nov 13;22(11):e22205 [FREE Full text] [doi: 10.2196/22205] [Medline: 33048825]
58. Basch CH, Zybert P, Reeves R, Basch CE. What do popular YouTube videos say about vaccines? Child Care Health Dev 2017 Jul;43(4):499-503. [doi: 10.1111/cch.12442] [Medline: 28105642]
59. Fadda M, Allam A, Schulz PJ. Arguments and sources on Italian online forums on childhood vaccinations: Results of a content analysis. Vaccine 2015 Dec 16;33(51):7152-7159. [doi: 10.1016/j.vaccine.2015.11.007] [Medline: 26592140]
60. Alessa A, Faezipour M. A review of influenza detection and prediction through social networking sites. Theor Biol Med Model 2018 Feb 01;15(1):2 [FREE Full text] [doi: 10.1186/s12976-017-0074-5] [Medline: 29386017]
61. Neumark Y, Flum L, Lopez-Quintero C, Shtarkshall R. Quality of online health information about oral contraceptives from Hebrew-language websites. Isr J Health Policy Res 2012 Sep 24;1(1):38 [FREE Full text] [doi: 10.1186/2045-4015-1-38] [Medline: 23006798]
62. Zarocostas J. How to fight an infodemic. The Lancet 2020 Feb;395(10225):676. [doi: 10.1016/s0140-6736(20)30461-x]
63. Eysenbach G. How to Fight an Infodemic: The Four Pillars of Infodemic Management. J Med Internet Res 2020 Jun 29;22(6):e21820 [FREE Full text] [doi: 10.2196/21820] [Medline: 32589589]
64. Siegrist M, Luchsinger L, Bearth A. The Impact of Trust and Risk Perception on the Acceptance of Measures to Reduce COVID-19 Cases. Risk Anal 2021 May;41(5):787-800 [FREE Full text] [doi: 10.1111/risa.13675] [Medline: 33438218]
65. Siegrist M, Zingg A. The Role of Public Trust During Pandemics. European Psychologist 2014 Jan 01;19(1):23-32. [doi: 10.1027/1016-9040/a000169]
66. Siegrist M. The influence of trust and perceptions of risks and benefits on the acceptance of gene technology. Risk Anal 2000 Apr;20(2):195-203. [doi: 10.1111/0272-4332.202020] [Medline: 10859780]
67. Slovic P. Trust, Emotion, Sex, Politics, and Science: Surveying the Risk-Assessment Battlefield. Risk Analysis 1999 Aug;19(4):689-701. [doi: 10.1111/j.1539-6924.1999.tb00439.x]
68. Agarwal V. A/H1N1 vaccine intentions in college students: an application of the theory of planned behavior. J Am Coll Health 2014;62(6):416-424. [doi: 10.1080/07448481.2014.917650] [Medline: 24779428]
69. de Perio MA, Wiegand DM, Evans SM. Low influenza vaccination rates among child care workers in the United States: assessing knowledge, attitudes, and behaviors. J Community Health 2012 Apr;37(2):272-281 [FREE Full text] [doi: 10.1007/s10900-011-9478-z] [Medline: 21938550]
70. Lau JT, Au DW, Tsui H, Choi K. Prevalence and determinants of influenza vaccination in the Hong Kong Chinese adult population. Am J Infect Control 2012 Sep;40(7):e225-e227. [doi: 10.1016/j.ajic.2012.01.036] [Medline: 22743046]
71. Matsui D, Shigeta M, Ozasa K, Kuriyama N, Watanabe I, Watanabe Y. Factors associated with influenza vaccination status of residents of a rural community in Japan. BMC Public Health 2011 Mar 04;11:149 [FREE Full text] [doi: 10.1186/1471-2458-11-149] [Medline: 21375758]

Abbreviations

OR: odds ratio

Edited by C Basch; submitted 23.07.21; peer-reviewed by A Benis, D Romer, C Yousef; comments to author 27.08.21; revised version received 03.10.21; accepted 18.11.21; published 06.01.22.

Please cite as:
Lee M, You M
Direct and Indirect Associations of Media Use With COVID-19 Vaccine Hesitancy in South Korea: Cross-sectional Web-Based Survey
J Med Internet Res 2022;24(1):e32329
URL: https://www.jmir.org/2022/1/e32329
doi:10.2196/32329
PMID:34870605

Original Paper

The Evolution and Disparities of Online Attitudes Toward COVID-19 Vaccines: Year-long Longitudinal and Cross-sectional Study

Chunyan Zhang[1*], MS; Songhua Xu[1*], PhD; Zongfang Li[1*], PhD; Ge Liu[2], BS; Duwei Dai[1], MS; Caixia Dong[1], MS

[1]Institute of Medical Artificial Intelligence, The Second Affiliated Hospital of Xi'an Jiaotong University, Xi'an, China

[2]School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China

[*]these authors contributed equally

Corresponding Author:
Songhua Xu, PhD
Institute of Medical Artificial Intelligence
The Second Affiliated Hospital of Xi'an Jiaotong University
No.157 Xiwu Road
Xi'an, 710004
China
Phone: 86 18710823698
Email: songhua_xu1@163.com

Abstract

Background: Due to the urgency caused by the COVID-19 pandemic worldwide, vaccine manufacturers have to shorten and parallel the development steps to accelerate COVID-19 vaccine production. Although all usual safety and efficacy monitoring mechanisms remain in place, varied attitudes toward the new vaccines have arisen among different population groups.

Objective: This study aimed to discern the evolution and disparities of attitudes toward COVID-19 vaccines among various population groups through the study of large-scale tweets spanning over a whole year.

Methods: We collected over 1.4 billion tweets from June 2020 to July 2021, which cover some critical phases concerning the development and inoculation of COVID-19 vaccines worldwide. We first developed a data mining model that incorporates a series of deep learning algorithms for inferring a range of individual characteristics, both in reality and in cyberspace, as well as sentiments and emotions expressed in tweets. We further conducted an observational study, including an overall analysis, a longitudinal study, and a cross-sectional study, to collectively explore the attitudes of major population groups.

Results: Our study derived 3 main findings. First, the whole population's attentiveness toward vaccines was strongly correlated (Pearson r=0.9512) with official COVID-19 statistics, including confirmed cases and deaths. Such attentiveness was also noticeably influenced by major vaccine-related events. Second, after the beginning of large-scale vaccine inoculation, the sentiments of all population groups stabilized, followed by a considerably pessimistic trend after June 2021. Third, attitude disparities toward vaccines existed among population groups defined by 8 different demographic characteristics. By crossing the 2 dimensions of attitude, we found that among population groups carrying low sentiments, some had high attentiveness ratios, such as males and individuals aged ≥40 years, while some had low attentiveness ratios, such as individuals aged ≤18 years, those with occupations of the 3rd category, those with account age <5 years, and those with follower number <500. These findings can be used as a guide in deciding who should be given more attention and what kinds of help to give to alleviate the concerns about vaccines.

Conclusions: This study tracked the year-long evolution of attitudes toward COVID-19 vaccines among various population groups defined by 8 demographic characteristics, through which significant disparities in attitudes along multiple dimensions were revealed. According to these findings, it is suggested that governments and public health organizations should provide targeted interventions to address different concerns, especially among males, older people, and other individuals with low levels of education, low awareness of news, low income, and light use of social media. Moreover, public health authorities may consider cooperating with Twitter users having high levels of social influence to promote the acceptance of COVID-19 vaccines among all population groups.

(J Med Internet Res 2022;24(1):e32394) doi:10.2196/32394

KEYWORDS

COVID-19; vaccine; attitude; Twitter; data mining; pandemic; population group; evolution; disparity

Introduction

Background

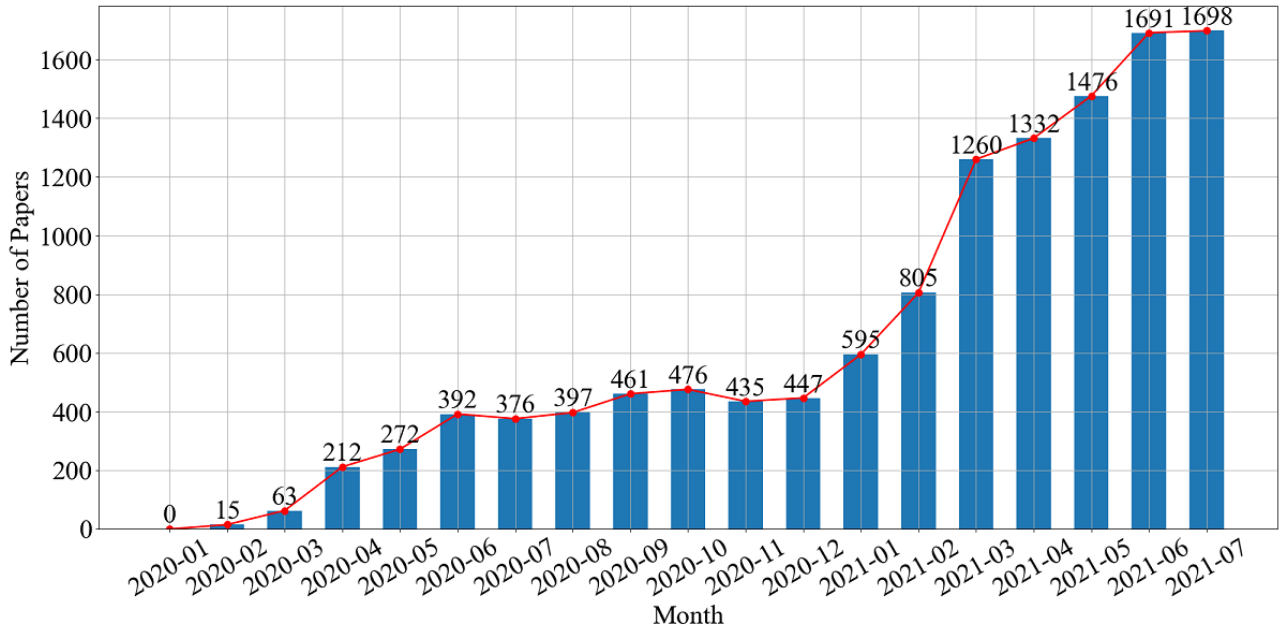
Since the emergence of the COVID-19 pandemic in 2019, human health and life have been gravely jeopardized globally. Governments and public health agencies worldwide primarily implemented the following 2 measures to control this pandemic: (1) nonpharmaceutical preventive methods, such as social distancing [1], and (2) COVID-19 vaccine development and mass vaccination to achieve herd immunity [2]. However, implementing nonpharmaceutical interventions is only a short-term solution since it will seriously affect the development of society. Vaccines, on the other hand, are more effective against infectious diseases.

Traditionally, developing a new vaccine from scratch is a complex process, which takes considerable time to accomplish. The main procedures of traditional vaccine development include preclinical studies (about 2-4 years); phase I, II, and III trials (about 5-7 years total); and manufacturing and approval (about 1-2 years) [3]. However, due to the great urgency of the COVID-19 pandemic, vaccine production was accelerated by shortening and paralleling the vaccine development steps. Although all usual safety and efficacy monitoring mechanisms were guaranteed to remain in place, varied attitudes toward these new vaccines have arisen among different population groups. Therefore, it is essential for us to study the evolution and disparities of attitudes across population groups accompanying the introduction of COVID-19 vaccines.

Literature Review

To identify COVID-19 vaccine-related literature, we searched the World Health Organization COVID-19 database [4] with the keywords “vaccine” and “vaccination.” This database is a comprehensive multilingual source of COVID-19 literature from various bibliographic databases (including MEDLINE, PubMed, and Scopus), hand searching, and the addition of other expert-referred scientific articles. Through filtering out preprints, choosing papers written in English, and excluding papers with incomplete information, a total of 12,403 papers were retrieved from the outbreak of the pandemic to July 31, 2021. Then, we counted and plotted the number of papers published each month, as shown in Figure 1. Since the first COVID-19 vaccine-related paper was published in February 2020, the number of papers grew fast until June 2020, and then remained steady from June to December 2020. After January 2021, it increased sharply again.

Figure 1. Monthly statistics on newly published papers related to COVID-19 vaccines.



In order to better understand the current research state of public attitudes toward COVID-19 vaccines, we filtered papers with the term “attitude” in the titles, and “cross-sectional” or “longitudinal” in the titles or abstracts, and retrieved 85 relevant papers. Then, we identified the data collecting methods used in these studies manually, and discovered that there were primarily 2 types as follows: survey (81 papers, 95%) and data mining (4 papers, 5%). Studies involving surveys mainly adopted a cross-sectional design to investigate public attitudes toward COVID-19 vaccination over a short period, while studies involving data mining employed either a longitudinal or cross-sectional design, but rarely both. We conducted a detailed literature review of COVID-19 vaccine-related studies involving these 2 frequently used analysis methods (cross-sectional and longitudinal).

A cross-sectional study analyzes data collected from a population or a predefined subset at a single point in time. Many studies have used this method, primarily through surveys, to explore populations’ attitudes toward receiving COVID-19

vaccination and the factors that affect these attitudes. For example, Lazarus et al [5] surveyed individuals randomly across 19 countries in June 2020. They concluded that the levels of willingness to accept a COVID-19 vaccine were insufficient to meet the requirements for community immunity in most of the 19 countries. Many cross-sectional studies discerned that demographic factors play an important role in vaccine acceptability [6-9]. Khubchandani [6] revealed that in the United States, the highest prevalence of COVID-19 vaccine hesitancy existed in some groups, such as African Americans, Hispanics, and individuals with lower education and incomes. Petravić et al [7] found that in Slovenia, a higher intention to get vaccinated was associated with men, older respondents, physicians, medical students, etc. In addition, some studies applied mining of social media data to perform cross-sectional analysis [10,11]. Hou et al [10] investigated vaccine-related posts from 5 global metropolises between June and July 2020. They discovered that vaccine hesitancy was prevalent worldwide and negative tweets attracted higher engagement on social media. While the cross-sectional method helps us identify diverse attitudes toward vaccines among different demographic groups at a particular time, it cannot track the evolution of attitudes over time.

A longitudinal study is a method that observes some specific variables over an extended period of time. Many studies have applied this method to track trends in population attitudes toward COVID-19 vaccines based on data mining, as data mining can process long-term and large-scale data. Pullan et al [12] and An et al [13] analyzed data from Google Trends, and they both found that the number of searches related to COVID-19 vaccines had increased during the pandemic. Yin et al [14] proposed a novel behavioral dynamics model on Weibo messages to analyze vaccine acceptance in China, and they demonstrated that Chinese individuals were inclined to be positive about side effects over time. Furthermore, many studies explored tweets related to COVID-19 vaccines to understand the evolution of public concerns and sentiments in different regions [15-17]. These studies mainly revealed that public concerns and sentiments on COVID-19 vaccines fluctuated with time and geography, and had strong correlations with some major events about COVID-19 vaccines. Although the longitudinal method applied in the above studies can track the general attitude trend of a population, it cannot identify the disparities of attitudes among demographic groups.

Considering the above-mentioned limitations, we combined cross-sectional and longitudinal analyses in this work to study the online attitudes toward COVID-19 vaccines based on data mining results of tweets. By doing this, we can not only track long-term evolution, but also discern the disparities of attitudes among various population groups. In addition, this work explored the correlation between the whole population's attentiveness toward vaccines and official COVID-19 statistics, and analyzed the abrupt influences of some major vaccine-related events. These findings can be used as a guide to assist governments and public health organizations in monitoring the trends of different population groups and relieving the low sentiments of specific groups. It is worth mentioning that the method proposed in this study can be easily reutilized to track the attitude evolution of population groups toward any other public health events. The source code developed in this study has been publicly released via GitHub for follow-up research [18].

The remainder of this paper is organized as follows. The Methods section first introduces the data collection and preprocessing procedures, and then presents the structure of a 2-step methodology with its essential design details. The Results section analyzes the mining outcomes from multiple dimensions. Finally, the Discussion section concludes the work.

Methods

Data Collection and Preprocessing

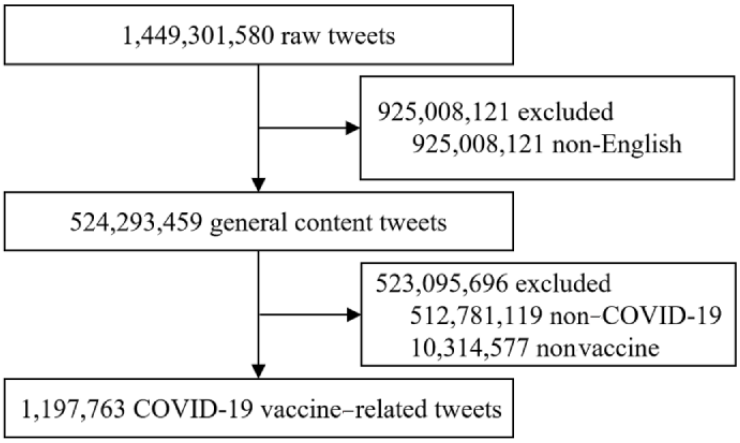
The Twitter data used in this study were randomly collected with our self-designed program using the Twitter application programming interface (API) [19] from June 9, 2020, to July 31, 2021, which covered some critical phases concerning the development and inoculation of COVID-19 vaccines. Moreover, the detailed user metadata (such as user name, biography, profile image, user creation time, and follower number) of each tweet were collected simultaneously by setting the Twitter API's optional query parameters. In total, over 1.4 billion tweets were collected during the research period.

So far, there have been a number of publicly released COVID-19 data sets available for scientific research, such as the data set released by Chen et al [20]. Compared with these public COVID-19 data sets, our data set has the following 2 obvious advantages for our research. First, public COVID-19 data sets only contain tweet IDs in compliance with Twitter's terms and conditions [21], so extra effort is needed to hydrate all tweet contents from tweet IDs [22]. In contrast, our data set needs less processes since all tweet contents are immediately available without hydration. The second advantage is much more crucial and ultimately led us to decide to use our data set. For public COVID-19 data sets, only up-to-date details on users can be retrieved from user IDs (extracted from hydrated tweet objects) at the time of hydration [23], while our data set already has detailed user metadata at the posting time of the tweets. Usually, user metadata would change more or less over 1 year, so the data at the time of hydration may not be able to infer the demographic characteristics of Twitter users at the posting time of the tweets. Therefore, our data set is more suitable than public COVID-19 data sets for studying the attitudes under various demographic characteristics.

As each tweet in our data set contains a detailed user profile, tweet text, a creation time, a location, statistics, and some other structured data, it can be treated as one online participant with the characteristics of an individual or organization, carrying an attitude for some specific topic. Since English is the most widely spoken language worldwide, we first excluded 925,008,121 non-English tweets by the language attribute of the tweet object, and obtained 524,293,459 English tweets on general content (hereinafter referred to as general content tweets). Then, we excluded 512,781,119 non-COVID-19–related tweets with a filtering pattern composed of 590 COVID-19 keywords and hashtags according to Twitter COVID-19 filtering rules [24]. To concentrate on vaccine-related tweets, we further excluded

10,314,577 non-vaccine tweets. Finally, in this study, we mainly focused on 1,197,763 vaccine-related tweets during COVID-19 (Figure 2) and used general content tweets as a benchmark of general population distribution.

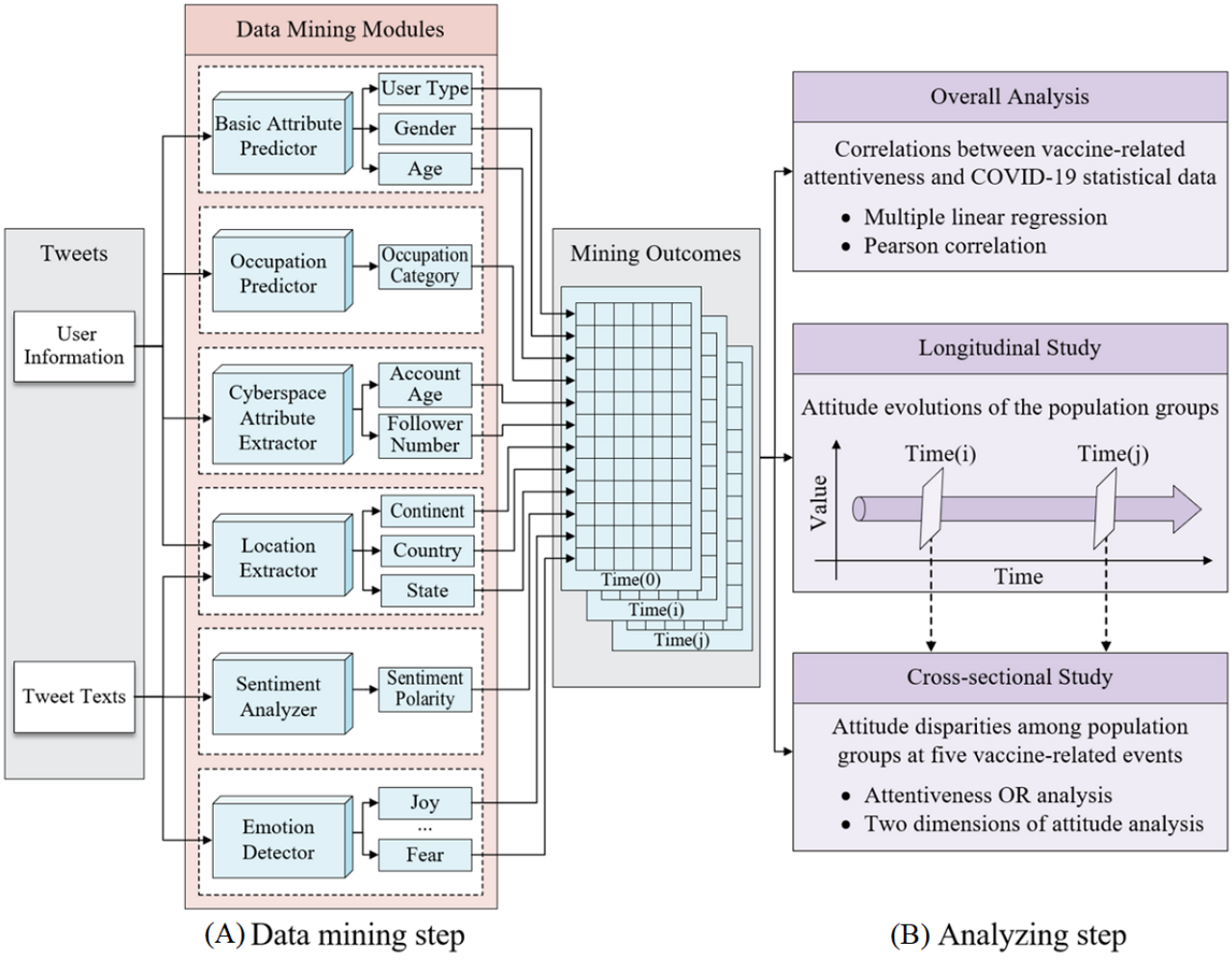
Figure 2. Data selection.



Study Design

We designed a 2-step methodology for this study, as shown in Figure 3. The implementation details of each step are described in the following sections.

Figure 3. The structure of the 2-step methodology in our study.



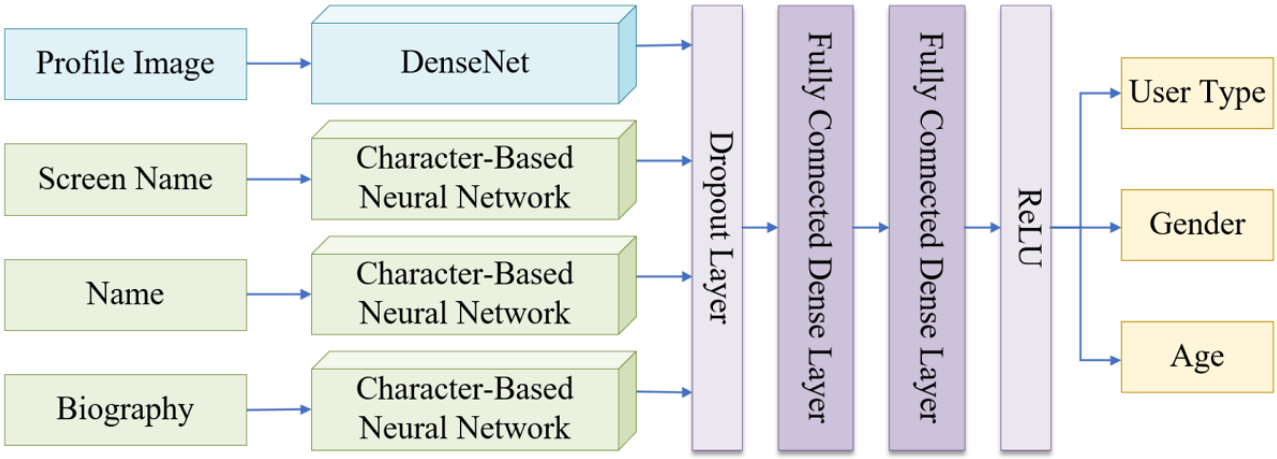
Data Mining Step

The data mining step plays a fundamental and decisive role in the entire research. We applied natural language processing, image processing, and tag extracting algorithms on tweets to extract users' real-world characteristics (user type, gender, age, occupation, and location), cyberspace characteristics (account age and follower number), sentiment polarities, and emotion types. One set of mining outcomes from a tweet constitutes one record of a user on a specific date. This step is analogous to the process of conventional questionnaire design and result collection. However, the data mining method can flexibly adjust the demographic characteristics that need to be analyzed, and acquire a stable amount of historical data from any population group during a long time period. This step contains 6 intelligent modules, which are described as follows.

Basic Attribute Predictor

This predictor, implemented with an open-source package of the M3 (multimodal, multilingual, and multiattribute) model [25], was employed to predict the probabilities of the following 3 basic demographic attributes: user type, gender, and age, through profile images, screen names, names, and biographies. As shown in Figure 4, the M3 model consists of 1 DenseNet module to process images, 3 character-based neural networks to process text, and finally 2 fully connected dense layers to predict the user type, gender, and age attributes. User type (individual or organization) and gender (male or female) are modeled as binary classification tasks, while age is modeled as a 4-class classification task with the following age groups: ≤18, 19-29, 30-39, and ≥40 years. This model was trained on a massive data set, including Twitter, IMDB, and Wikipedia data [26], and was fine-tuned to capture accurate demographic features. In a previous study, we had tested the M3 model on a subset of our tweet data set, and obtained the benchmark performance as follows: for user type, gender, and age attributes, the accuracy scores were 99.07%, 95.88%, and 77.65%, respectively, and the macro-F1 scores were 0.9860, 0.9572, and 0.7311, respectively [27]. Detailed information about this model can be found in a previous report [25].

Figure 4. The structure of the M3 (multimodal, multilingual, and multiattribute) model. DenseNet: dense convolutional network; ReLU: rectified linear unit.



Occupation Predictor

We constructed a word-based convolutional neural network for occupation inference through user information (such as biographies) and tweet text. As shown in Figure 5, this deep learning model consists of 1 encoding layer, 1 embedding layer, 3 parallel convolution blocks, and finally 1 dense layer to classify the results into 3 classes. Every convolution block is composed of a 1D convolution layer with a different kernel size (eg, 2, 3, and 4), a 1D max-pooling layer, and a flatten layer. The training data we used are from a publicly available data set [28] that has 5191 Twitter users annotated with 9-class occupation categories (OCs), which are defined based on the Standard Occupation Classification (SOC) from the United Kingdom [29]. However, the 9-class classification of Twitter occupation is a challenging task, and the results of the latest studies are not accurate enough for our study. For example, Pan et al [30] employed a 3-layer graph convolutional network (GCN) model to predict the 9-class occupations, and the best performance of accuracy was 61.0%. Therefore, we simplified the 9-class classification into a 3-class classification, and the OCs were abbreviated as OC1, OC2, and OC3. Considering the relationship of the SOC classes and the balance of the 3 categories in the training set, we designed a new occupation division, as shown in Table 1. By adopting 10-fold cross-validation, the accuracy of the occupation predictor reached 74.08% on the new data set.

Figure 5. The structure of the occupation predictor. Conv Block: convolution block; 1D-Conv(kernel_size=k): 1D convolution layer with a kernel size of k; 1D-MaxPool: 1D max-pooling layer; Flatten: flatten layer; OCi: the ith occupation category, i∈(1,3).

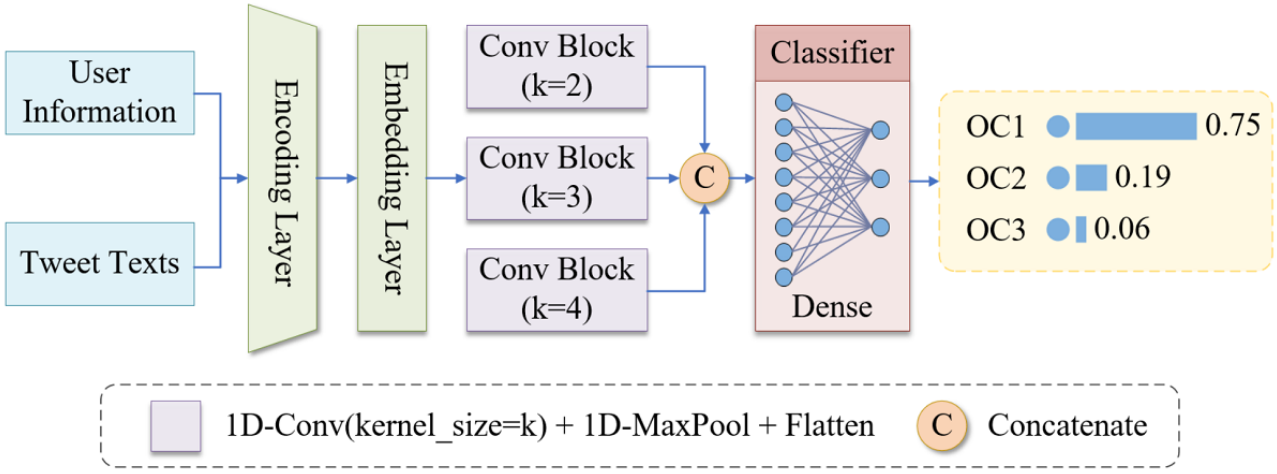


Table 1. The new occupation categories in our study and the original occupation categories in the Standard Occupation Classification hierarchy.

New occupation category (OC)	Original occupation category
OC1	C2: professional occupations
OC2	C1: managers, directors, and senior officials
	C3: associate professional and technical occupations
OC3	C4: administrative and secretarial occupations
	C5: skilled trade occupations
	C6: caring, leisure, and other service occupations
	C7: sales and customer service occupations
	C8: process, plant, and machine operatives
	C9: elementary occupations

Cyberspace Attribute Extractor

In this study, we only focused on the following 2 key attributes in cyberspace: account age and follower number. As a matter of fact, there are plenty of cyberspace attributes recorded in the tweet object, such as verified status and tweet number. The reason why we chose these 2 attributes is that account age can reflect the internet age of a user, and follower number can indicate a user's influence and usage level of social media to a certain extent. Additionally, Lyu et al [31] studied the characterization of population groups with varied attitudes toward COVID-19 vaccines and concluded that account age and follower number also have an influence on population attitudes toward vaccines. The account age of a user was calculated by subtracting the user's creation time from the tweet's creation time, and the follower number was retrieved directly from the "stats" field of the tweet object.

Location Extractor

We used the "geo" field in the tweet object and the "location" field in the profile of a Twitter user to efficiently extract location information, including the continent, country, and state. The extraction process was implemented in 2 steps. First, the location extractor called the Twitter API to query a place by the geocode in the "geo" field [32] and then obtained the exact location from the retrieved data. Second, if a tweet did not contain a geocode, the extractor used the "location" field to fuzzy inquire the location, which was implemented by GeoPy [33] and pycountry [34]. Through these 2 steps, approximately 63.41% of Twitter users' locations could be extracted.

Sentiment Analyzer

The sentiment analyzer was implemented with an open-source tool named Valence Aware Dictionary and Emotional Reasoner (VADER) [35]. It is a lexicon and rule-based sentiment analysis tool specifically attuned to sentiments expressed in social media. In this study, we used VADER to calculate the sentiment polarities (−1 to 1) of the tweet text and then divided the results into the following 3 subranges: negative (–1 to –0.05), neutral (–0.05 to 0.05), and positive (0.05 to 1).

Emotion Detector

The emotion detector was based on an open-source emotion recognition algorithm [36] on Twitter that utilized a character-based trained recurrent neural network algorithm. In the original study, it implemented 3 different emotion models. We selected the model of Ekman's 6 basic emotions [37] (anger, disgust, fear, joy, sadness, and surprise) to predict the emotion types of the tweets in this study.

Analyzing Step

Based on the mining outcomes, we conducted multiple analyses, including overall analysis, a longitudinal study, and a cross-sectional study, to detect the evolution and disparities of

attitudes toward COVID-19 vaccines among population groups during the study period. The multiple analyses are shown in Figure 3.

Concretely, in the overall analysis, multiple linear regression and Pearson correlation analysis were used to detect the impacts of official COVID-19 statistics, including confirmed cases, deaths, vaccinations, and reproduction rate, and some major vaccine-related events on the whole population's attentiveness toward vaccines. In the longitudinal study, to detect the evolution of attitudes among different population groups over time, a series of longitudinal contrasts with different demographic characteristics were displayed and analyzed. To further reveal the attitude patterns among population groups, a cross-sectional study, incorporating the benchmark of the population distribution of the general content tweets, was conducted at 5 vaccine-related events selected from the overall analysis.

The strength of the correlation and similarity in this paper using the guide that Evans [38] suggested for the absolute value of Pearson *r* can be described as follows: very weak (0 to 0.19), weak (0.20 to 0.39), moderate (0.40 to 0.59), strong (0.60 to 0.79), and very strong (0.80 to 1).

Results

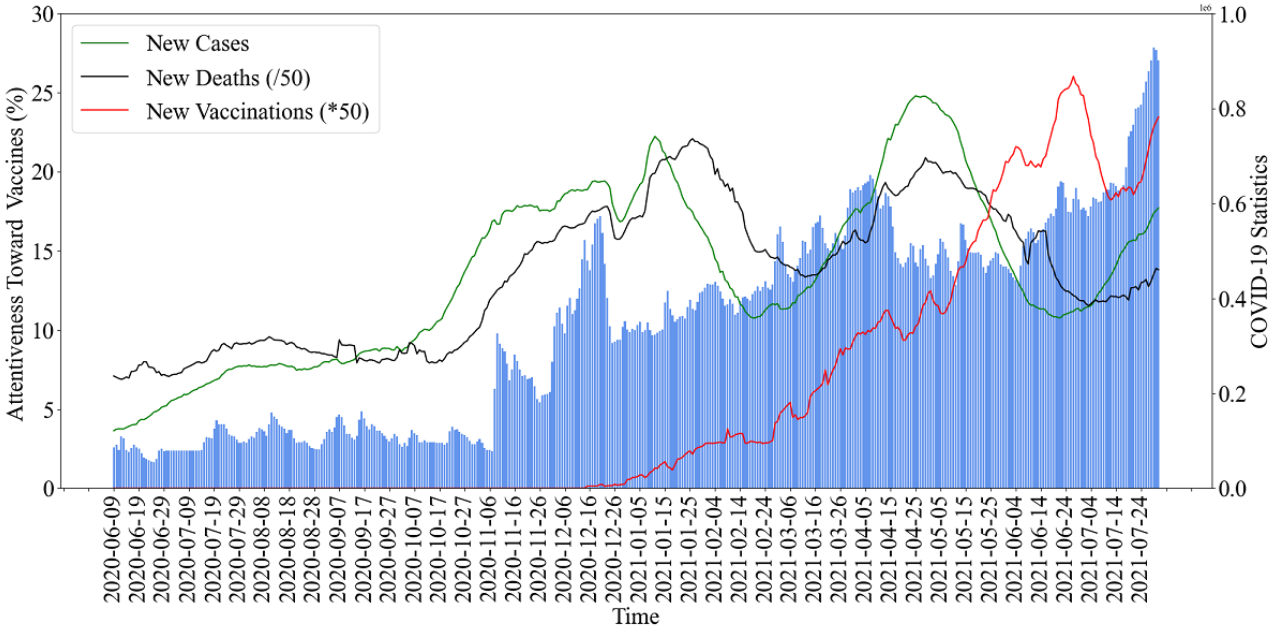
Overall Analysis

In this section, we explore the possible influencing factors of the whole population's attentiveness toward vaccines in 2 steps. First, studying the correlation between attentiveness and official COVID-19 statistics. Second, discovering the abrupt influences of some major vaccine-related events during COVID-19. At the end of this section, the data mining results of vaccine-related tweets and general content tweets have been displayed and analyzed in general.

The Influencing Factors of the Attentiveness Toward Vaccines

In order to eliminate possible fluctuations in the quantities of tweets captured daily, we used the percentage of vaccine-related tweets in COVID-19 tweets to represent the attentiveness toward vaccines (sometimes referred to as attentiveness for short in the following text) during COVID-19 in this study. Meanwhile, we obtained the global data of COVID-19 statistics from Our World in Data [39], including the daily numbers of new cases, new deaths, new vaccinations, total cases, total deaths, total vaccinations, and people vaccinated, as well as the reproduction rate. Three of the statistics are plotted as examples in Figure 6.

Figure 6. The whole population's attentiveness toward vaccines, and the COVID-19 statistics during the study period. The y-axis on the left is the level of attentiveness, and the y-axis on the right represents the numbers of COVID-19 statistics, which adopt different scales.



Since some variables of the COVID-19 statistics do not exert an immediate influence on the attentiveness toward vaccines, we applied Pearson correlation analysis on different time delays ($0 \leq lag \leq 30$ days) of attentiveness with each of the statistical variables, and found that the optimal lag at which Pearson *r* reached the absolute maximum value was different for each variable, as shown in Table 2. Moreover, all variables of the COVID-19 statistics showed positive correlations with attentiveness, except for reproduction rate, which showed a negative correlation, with $r=-0.4562$ at lag=10.

Table 2. Pearson correlation coefficients between COVID-19 statistics and attentiveness toward vaccines with lags that have absolute maximum r values within 30 days.

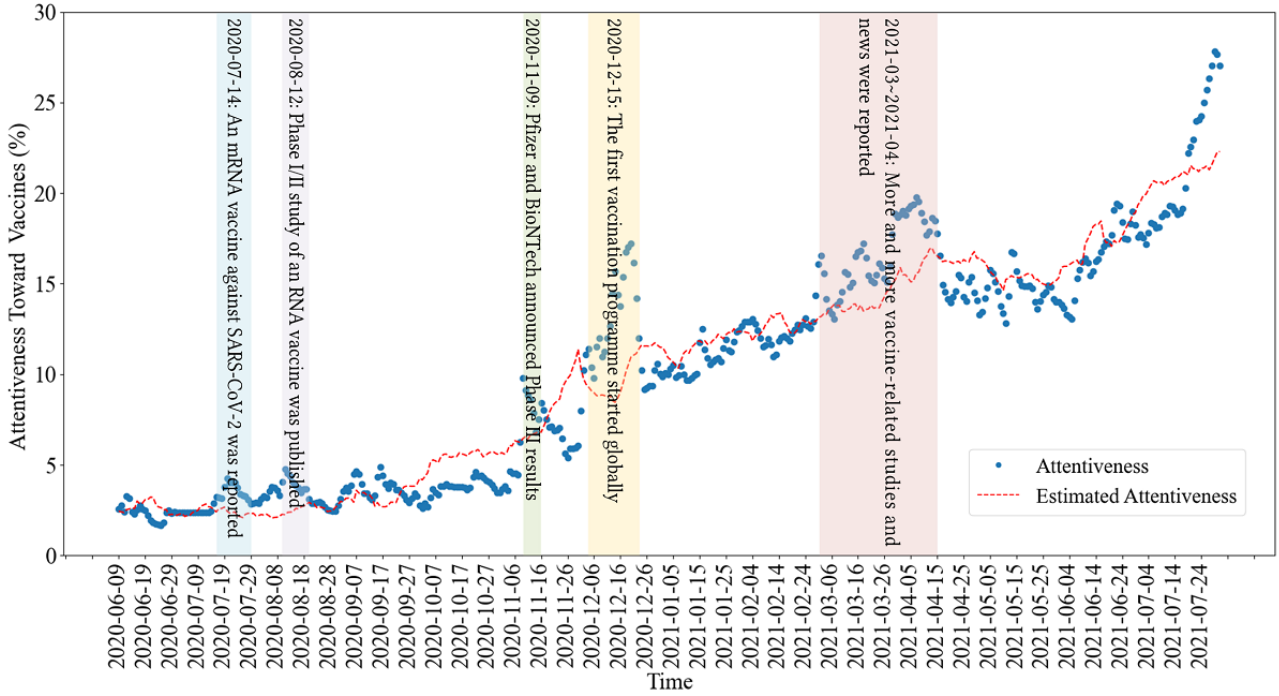
Variable	Lag days	Pearson *r*	*P* value
New cases	5	0.5917	<.001
New deaths	3	0.6543	<.001
New vaccinations	4	0.7843	<.001
Total cases	0	0.9093	<.001
Total deaths	0	0.9066	<.001
Total vaccinations	0	0.7433	<.001
People vaccinated	0	0.7364	<.001
Reproduction rate	10	−0.4562	<.001

Then, we took the 8 variables of COVID-19 statistics as independent variables (denoted as $X_i(t)$, $i \in (1,8)$ and $X_0(t)=1$, so $X(t)$ is a 9-dimensional vector) and attentiveness as a dependent variable (denoted as $Y(t)$). Multiple linear regression was used to analyze the relationship between $X(t)$ and $Y(t)$, which is expressed as follows:

In this formula, β is a coefficient vector of the shifted independent variables. Shift is a function to make a shift −L on $X(t)$. **L** is a shift vector composed of the lag values taken from Table 2, that is, L=(0,5,3,4,0,0,0,0,10). Since the shifting direction of these lags is for attentiveness, when shifting the independent variables $X(t)$, we take the opposite direction of **L**, that is **−L**.

By using the above regression model between the attentiveness $Y(t)$ and the COVID-19 statistical vector $X(t)$, we obtained the coefficient vector β, the adjusted R-squared value, which reached 0.9034, and the estimated attentiveness $\hat{Y}(t)$. The Pearson correlation between $Y(t)$ and $\hat{Y}(t)$ was 0.9512 (P<.001), which was higher than that for $Y(t)$ with any single independent variable $X_i(t)$ in Table 2. Therefore, the whole population's attentiveness toward vaccines and the estimated attentiveness by COVID-19 statistics showed a very strong positive correlation. The curves of the attentiveness $Y(t)$ and estimated attentiveness $\hat{Y}(t)$ are shown in Figure 7.

Figure 7. The whole population's attentiveness toward vaccines, estimated attentiveness by COVID-19 statistics, and labels of 5 major vaccine-related events.



Furthermore, we noticed that some abrupt jumps appeared in the attentiveness curve. An earlier study by Chen et al [20] concluded that Twitter discourse statistics can reflect major events at the time. Inspired by this, we calculated the periods when the attentiveness was significantly higher than the fitted curve and lasted more than 5 days. We used the standard deviation of the gaps between daily attentiveness and the fitted data as a threshold for judging the significant level. After that, we found that these periods exactly corresponded to some major vaccine-related events, as shown in Table 3. Compared to the

development process of COVID-19 vaccines, the time points of these vaccine-related events can roughly fall into the following phases: t_1, t_2, and t_3 in phase I/II/III of the vaccine clinical trial period, and t_4 and t_5 in the vaccination period. Since each manufacturer has its own schedule for developing COVID-19 vaccines, the description of the vaccine phases here is used only as a reference for this study.

Table 3. Some major events related to COVID-19 vaccines.

Time	Major event	Highest attentiveness percentage	Lasting days
July 14, 2020 (t_1)	"An mRNA vaccine against SARS-CoV-2—preliminary report" [40] was published.	4.27%	14
August 12, 2020 (t_2)	"Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults" [41] was published.	4.78%	12
November 9, 2020 (t_3)	Pfizer and BioNTech announced phase III results [42].	9.78%	6
December 15, 2020 (t_4)	The first mass vaccination program started globally [43].	17.48%	21
April 10, 2021 (t_5) selected from March to April, 2021	More and more vaccine-related studies and news were reported (see Multimedia Appendix 1).	18.92%	46

The above data analysis shows that the total online population's attentiveness toward COVID-19 vaccines was significantly correlated with COVID-19 statistics, including confirmed cases, deaths, vaccinations, and reproduction rate. Besides, attentiveness was also influenced by some vaccine-related events.

Statistics of Inferred Latent Characteristics

Two data mining experiments were conducted on 2 types of tweets as a comparison to infer different latent characteristics. The data mining methods are described in the Data Mining Step of the Methods section. The first experiment was performed on 1,197,763 vaccine-related tweets covering 1 year, and the second experiment was on 100,000 general content tweets selected from 5 major vaccine-related events (20,000 samples at each event) described in Table 3. The purpose of the second experiment was to get a benchmark of the general population distribution. The overall statistics of the 2 types of tweets are summarized and shown in Table 4. *P* values are calculated by the chi-square test.

Table 4. Overall statistics of vaccine-related tweets and general content tweets.

Characteristic	Vaccine-related tweets (N=1,197,763)	General content tweets at 5 major time points (N=100,000)	*P* value
User type, n (%)			<.001
Individual	1,079,105 (90.09)	94,560 (94.56)	
Organization	118,658 (9.91)	5440 (5.44)	
Gender, n (%)			<.001
Male	661,511 (61.30)	50,156 (53.04)	
Female	417,594 (38.70)	44,404 (46.96)	
Age (years), n (%)			<.001
≤18	157,395 (14.59)	35,036 (37.05)	
19-29	254,920 (23.62)	32,142 (33.99)	
30-39	202,451 (18.76)	11,112 (11.75)	
≥40	464,339 (43.03)	16,270 (17.21)	
Occupation category (OC), n (%)			<.001
OC1	385,276 (35.70)	20,102 (21.26)	
OC2	347,257 (32.18)	31,176 (32.97)	
OC3	346,572 (32.12)	43,282 (45.77)	
Location–continent[a], n (%)			<.001
North America	274,565 (40.12)	21,202 (37.45)	
Europe	179,683 (26.26)	13,132 (23.20)	
Asia	106,713 (15.60)	12,388 (21.88)	
Africa	67,805 (9.91)	7320 (12.93)	
Oceania	39,414 (5.76)	1250 (2.21)	
South America	15,794 (2.31)	1292 (2.28)	
Antarctica	301 (0.04)	26 (0.05)	
Location–country[a,b], n (%)			<.001
United States	214,606 (31.36)	18,163 (32.09)	
United Kingdom	106,337 (15.54)	6334 (11.19)	
India	51,698 (7.56)	2969 (5.24)	
Canada	42,554 (6.22)	1538 (2.72)	
Australia	20,477 (2.99)	474 (0.84)	
Account age (years), n (%)			<.001
<5	478,914 (44.38)	58,341 (61.70)	
5-10	349,595 (32.40)	26,333 (27.85)	
≥10	250,596 (23.22)	9886 (10.45)	
Follower number, n (%)			<.001
<500	619,808 (57.44)	55,745 (58.95)	
500-5000	368,780 (34.17)	32,742 (34.63)	
≥5000	90,517 (8.39)	6073 (6.42)	
Sentiment polarity			<.001
Overall (–1 to 1), mean (SD)	0.0161 (0.4591)	0.1215 (0.4566)	<.001
Negative (–1 to −0.05), n (%)	411,990 (38.18)	41,555 (43.95)	
Neutral (–0.05 to 0.05), n (%)	292,273 (27.08)	29,987 (31.71)	

Characteristic	Vaccine-related tweets (N=1,197,763)	General content tweets at 5 major time points (N=100,000)	*P* value
Positive (0.05 to 1), n (%)	374,842 (34.74)	23,018 (24.34)	
Emotion, n (%)			<.001
Fear	528,667 (48.99)	21,703 (22.95)	
Joy	313,423 (29.04)	31,819 (33.65)	
Surprise	153,807 (14.25)	28,690 (30.34)	
Sadness	42,124 (3.90)	8940 (9.45)	
Anger	23,814 (2.21)	2570 (2.72)	
Disgust	17,270 (1.60)	838 (0.89)	

[a]Under location (continent and country) characteristics, the total number of vaccine-related tweets with location information was 684,275, and the total number of general content tweets with location information was 56,610.

[b]The top 5 countries with the most vaccine-related tweets are selected.

From Table 4, it can be seen that there were significant differences (*P*<.001) between vaccine-related tweets and general content tweets. For example, the proportions for males and females were 53.04% and 46.96% in general content tweets, which were relatively balanced in gender. However, the gender gap enlarged in vaccine-related tweets to 22.60% (61.30% for males and 38.70% for females). The top 3 emotions in general content tweets were *joy* (33.65%), *surprise* (30.34%), and *fear* (22.95%), while in vaccine-related tweets, the top 3 were *fear* (48.99%), *joy* (29.04%), and *surprise* (14.25%).

In the next 2 sections, we further analyzed these long-term and multicharacteristic data in longitudinal and cross-sectional studies.

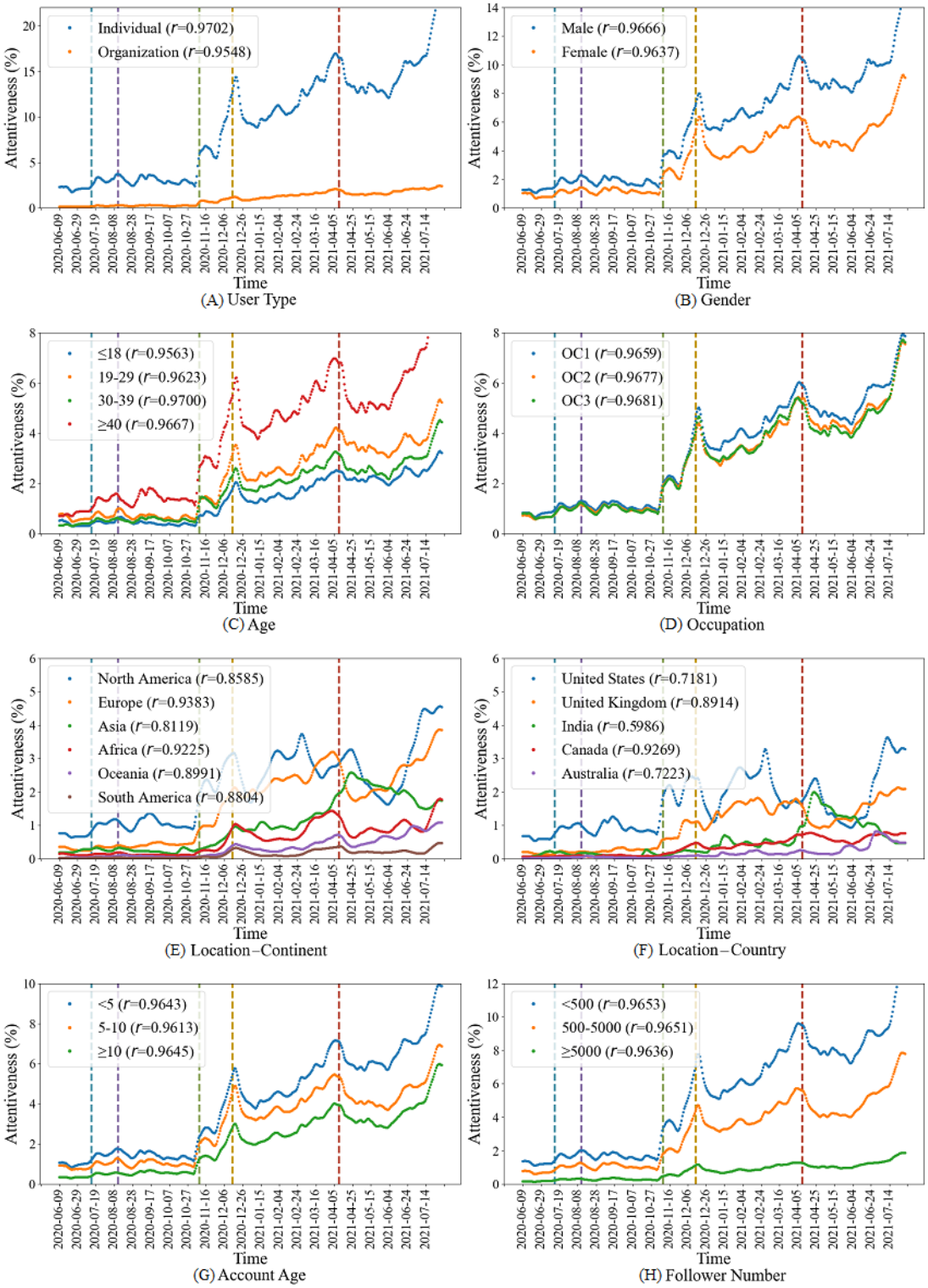
Longitudinal Study

In this section, we analyzed the attitude evolution of population groups from the following 2 aspects: attentiveness and sentiments toward vaccines.

Attentiveness Toward Vaccines Among Different Population Groups

Based on the data mining outcomes, we obtained the daily attentiveness toward vaccines of population groups by calculating the percentages of vaccine-related tweets in COVID-19 tweets with different characteristics. As shown in Figure 8, in general, the attentiveness of each group increased over time. At the 5 major vaccine-related events, most of the population groups had local peaks, similar to the whole population's attentiveness. By calculating the Pearson *r* values, it was shown that the attentiveness of all population groups had very strong similarities ($r>0.80$) with that of the whole population, except for the United States and Australia that had strong similarities ($0.60<r<0.79$) and India that had a moderate similarity ($0.40<r<0.59$).

Figure 8. Attentiveness toward vaccines of different population groups. The vertical dashed lines represent the time points of 5 major vaccine-related events as follows: July 14, 2020 (*t*1), August 12, 2020 (*t*2), November 9, 2020 (*t*3), December 15, 2020 (*t*4), and April 10, 2021 (*t*5). The *r* values in the legend boxes are the Pearson correlation coefficients between each population group and the whole population. OC: occupation category.



Furthermore, except for 2 location characteristics, population groups under the 6 demographic characteristics exhibited consistent differences throughout the study period. For example, males always had a higher level of attentiveness than females. In contrast, there was no specific pattern for population groups under the 2 location characteristics. In particular, North America almost had the highest attentiveness among all continents during the pandemic, while Europe and Asia surpassed it in some periods. Moreover, the United States almost had the highest attentiveness among all countries, while the United Kingdom and India surpassed it in some periods.

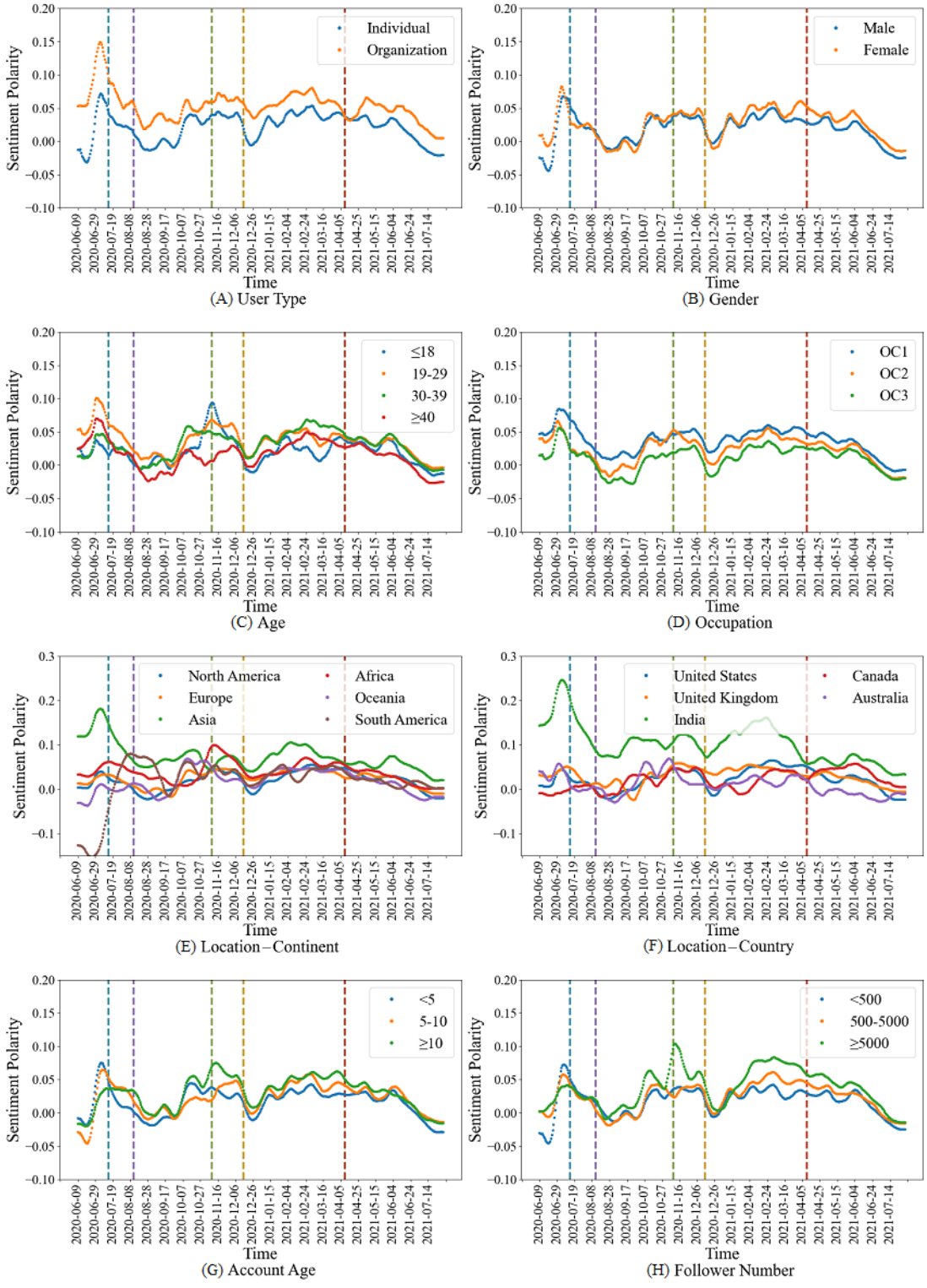
In summary, most of the population groups had very strong similarities with the whole population regarding attentiveness, that is, attentiveness increased over time with some local peaks,

indicating that they might be affected by the vaccine-related events as well. Moreover, there existed consistent group differences in the evolution of attentiveness under the demographic characteristics, except for the 2 location characteristics.

Sentiments Toward Vaccines Among Different Population Groups

As shown in Figure 9, we calculated the daily mean sentiment polarities of the population groups under the demographic characteristics. In general, the sentiments of all population groups fluctuated greatly in the early period of the vaccine development, and gradually stabilized at the start of vaccination (t_4). Nevertheless, they all went down after June 2021.

Figure 9. Sentiments toward vaccines of different population groups. OC: occupation category.



In particular, the sentiments of organizations were more positive than individuals. Females were sometimes a bit less positive than males in the development phase of vaccines, but were more positive than males after the inoculation started. Among the 4 age groups, the sentiments of people aged ≥40 and ≤18 years were almost the lowest during the study period. Among the 3 categories of occupations, the sentiments of OC1 were the highest, while those of OC3 were always the lowest. Under the 2 location characteristics, South America exhibited a different sentiment trend compared with other continents. Asia among the continents and India among the countries had the highest sentiments, and both showed downward trends. Among the 3 account age groups, the group of <5 years almost had the lowest sentiments. Among the follower number groups, the group of <500 followers nearly had the lowest sentiments, while the group of ≥5000 followers had the highest sentiments, except for the period before t_2.

In summary, the sentiments differed among population groups and fluctuated a lot in the early period of vaccine development, which suggested that different populations might hold different and immature views at the beginning. After June 2021, there were downward trends in all populations, indicating that populations might become less positive toward vaccines than before.

Cross-sectional Study

In the previous section, we mainly focused on the long-term evolution of population attitudes toward vaccines among COVID-19–related tweets, ignoring the general population distribution. Actually, the sizes of population groups vary greatly with respect to population characteristics. Thus, it is meaningful and essential to investigate the attentiveness ratios toward vaccines among different population groups under the benchmark of the general population distribution. Therefore, in this section, we conducted 5 cross-sectional analyses at 5 major vaccine-related events, by applying the odds ratio (OR) to represent the attentiveness ratio toward vaccines of each population group. Due to the complexity of attentiveness among continents and countries under the 2 location characteristics, we only analyzed the 6 demographic characteristics.

As shown in Table 5, considering the general population distribution, the attentiveness ratios toward vaccines of organizations were higher than that of individuals at all the time points, with the OR ranging from 1.44 (95% CI 1.28-1.61) to 2.01 (95% CI 1.70-2.39). The attentiveness ratios of females were lower than that of males, with the OR ranging from 0.61 (95% CI 0.57-0.66) to 0.82 (95% CI 0.77-0.88). Under the age characteristic, the OR increased progressively with age at all 5 time points. The ≥40 age group always had the highest attentiveness ratios, with the OR ranging from 3.70 (95% CI 3.53-3.87) to 8.81 (95% CI 8.53-9.10). Under the occupation characteristic, OC3's attentiveness ratios were the lowest among the 3 categories of occupations, with the OR ranging from 0.37 (95% CI 0.30-0.44) to 0.44 (95% CI 0.39-0.49). The attentiveness ratios of the account age groups showed the same trends as the real-world age groups, with the attentiveness OR increasing with account age. Under the follower number characteristic, a higher number of followers was associated with a higher attentiveness OR.

Table 5. Attentiveness odds ratios toward vaccines at 5 major vaccine-related events.

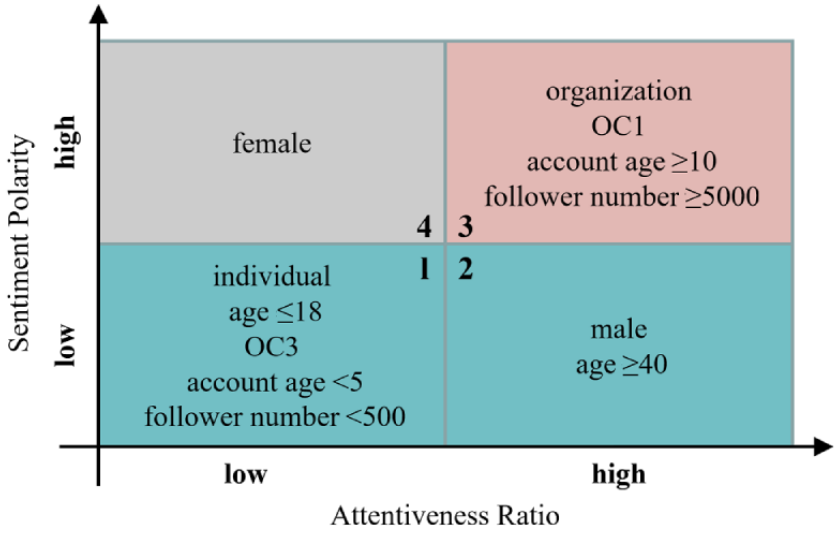
Characteristic	Major vaccine-related events				
	t_1: July 14, 2020, attentiveness OR[a] (95% CI)	t_2: August 12, 2020, attentiveness OR (95% CI)	t_3: November 9, 2020, attentiveness OR (95% CI)	t_4: December 15, 2020, attentiveness OR (95% CI)	t_5: April 10, 2021, attentiveness OR (95% CI)
User type					
Individual	1 (ref[b])	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Organization	1.51 (1.20-1.89)	1.46 (1.22-1.76)	2.01 (1.70-2.39)	1.44 (1.28-1.61)	1.98 (1.79-2.19)
Gender					
Male	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
Female	0.75 (0.66-0.86)	0.79 (0.71-0.87)	0.66 (0.60-0.73)	0.82 (0.77-0.88)	0.61 (0.57-0.66)
Age (years)					
≤18	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
19-29	1.68 (1.55-1.82)	1.49 (1.42-1.56)	1.85 (1.78-1.93)	2.31 (2.23-2.39)	2.13 (2.05-2.20)
30-39	3.55 (3.26-3.86)	2.31 (2.19-2.44)	4.48 (4.30-4.66)	4.39 (4.23-4.55)	5.22 (5.03-5.42)
≥40	5.26 (4.89-5.67)	3.70 (3.53-3.87)	7.34 (7.08-7.60)	7.12 (6.90-7.35)	8.81 (8.53-9.10)
Occupation category (OC)					
OC1	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
OC2	0.52 (0.44-0.63)	0.62 (0.55-0.70)	0.66 (0.61-0.72)	0.52 (0.48-0.57)	0.59 (0.55-0.64)
OC3	0.37 (0.30-0.44)	0.44 (0.39-0.49)	0.42 (0.39-0.46)	0.41 (0.38-0.45)	0.42 (0.39-0.45)
Account age (years)					
<5	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
5-10	1.31 (1.10-1.55)	1.34 (1.20-1.49)	1.67 (1.54-1.81)	2.28 (2.12-2.46)	2.25 (2.08-2.44)
≥10	1.93 (1.54-2.40)	1.62 (1.40-1.87)	2.88 (2.62-3.17)	3.71 (3.40-4.04)	3.28 (3.01-3.58)
Follower number					
<500	1 (ref)	1 (ref)	1 (ref)	1 (ref)	1 (ref)
500-5000	0.97 (0.82-1.14)	1.02 (0.92-1.13)	1.04 (0.96-1.13)	1.07 (0.99-1.14)	1.02 (0.95-1.10)
≥5000	1.32 (1.00-1.74)	1.20 (1.00-1.44)	1.69 (1.48-1.92)	1.43 (1.27-1.61)	1.40 (1.23-1.59)

[a]OR: odds ratio.

[b]ref: reference.

To further discern the deep law of attitudes under different demographic characteristics after the inoculation started, we crossed the 2 dimensions of attitude (the attentiveness ratio and sentiment polarity) according to the cross-sectional results at t_4 and t_5. The levels of the attentiveness ratios and sentiment polarities can be found in Table 5 and Figure 9 (see Multimedia Appendix 2 for the specific sentiment polarities), respectively. Then, all the population groups were divided into 4 categories, as shown in Figure 10. Each grid in this figure is 1 category combining high or low levels of the 2 dimensions. For example, the 1st category is a combination of low attentiveness ratio and low sentiment polarity, and it includes individuals with age ≤18 years, OC3 (occupations of the 3rd category), account age <5 years, and follower number <500.

Figure 10. Four categories of population groups by crossing the 2 dimensions of attitude toward vaccines. OC: occupation category.



From Figure 10, it can be seen that the 1st category and 2nd category both had low sentiments toward vaccines, with either low or high levels of attentiveness ratios. The 1st category was mainly composed of people with relatively low levels of education, low awareness of news, low income, and light use of social media, while the 2nd category was composed of males and older people. The 3rd category had high sentiments and high attentiveness ratios, indicating that these population groups had persistent attention and trust in the COVID-19 vaccines. Some groups in the 3rd category were groups with relatively high levels of social influence, both in reality (organizations) and in cyberspace (account age ≥10 years, follower number ≥5000). These findings can provide comprehensive guidance for governments and public health organizations in deciding who should be given more attention and targeted help, and who can be considered to promote the publicity of vaccines. In the Discussion section, we will further investigate the possible reasons for the categories carrying low sentiments, and provide some practical suggestions for interventions accordingly.

Discussion

Principal Findings

In this study, we acquired and analyzed a year-long collection of tweets, from June 9, 2020, to July 31, 2021, to discover the evolution and disparities of attitudes toward COVID-19 vaccines among various online population groups. Overall, the whole population's attentiveness toward COVID-19 vaccines increased over time with some local fluctuations during the study period. This study demonstrated that this attentiveness had a very strong correlation (Pearson r=0.9512) with official COVID-19 statistics, such as confirmed cases and deaths, and it was noticeably influenced by some major vaccine-related events. Studies conducted on other online platforms, such as Google and Baidu, also had results similar to ours. For example, Hu et al [44] found that there was a correlation between the Google Trends of relative search volume for COVID-19 and the daily number of new cases. Besides, by comparing demographic composition and sentiments between vaccine-related tweets and general content tweets, we found that there were significant differences (P<.001). In particular, the whole population had lower sentiments and more *fear* toward vaccines than general topics.

By analyzing the attentiveness evolution toward vaccines under 8 demographic characteristics, we observed that, except for the United States, Australia, and India, all population groups exhibited very strong similarities with the whole population. As for the sentiment evolution toward vaccines, we found that different populations initially held different and fluctuated sentiments in the early stage of vaccine development, and then, the sentiments gradually stabilized and tended to relatively positive levels at the beginning of vaccination, but after June 2021, they all had a considerably downward trend. The research findings of Yan et al [45] and Hu et al [46] on sentiment toward COVID-19 vaccines are partially consistent with ours. However, their studies do not have data from June to July 2021 to verify the downward trend in sentiment discovered in our study. By reading news reports during this period, we noticed some possible clues for this trend. The most likely one is the spread of the Delta variant, which is more contagious than the other coronavirus strains [47]. It is claimed that vaccination is still the best protection against the Delta variant. Thus, the downward trend in sentiment is a critical warning signal for governments and public health agencies, calling for more attention dedicated to public concerns regarding COVID-19 vaccines.

Furthermore, there are significant attitude disparities toward COVID-19 vaccines across population groups. By crossing the 2 dimensions of attitude (the attentiveness ratio and sentiment polarity), we found that among population groups carrying low sentiments, some have low attentiveness ratios (the 1st category in Figure 10), such as individuals with age ≤18 years, occupations of the 3rd category, account age <5 years, and follower number <500, while some have high attentiveness ratios (the 2nd category in Figure 10), such as males and individuals with age ≥40 years.

We investigated and inferred the internal reasons for the low sentiments, and found some corresponding epidemiological studies that can confirm our inference. For the 1st category of the population, the low sentiments may be derived from the insufficient knowledge and distrust of vaccines based on

education status, news awareness, economic conditions, level of social media usage, etc. This finding appears consistent with the finding of Paul et al [48], who revealed that distrustful attitudes toward vaccination were higher among individuals with lower levels of education, lower annual income, poor knowledge of COVID-19, and poor compliance with government COVID-19 guidelines. The 2nd category of the population may be aware of the high risks of COVID-19 related to their own physical conditions, so they are prone to incur negative sentiments. Medical evidence indicates that males [49] compared to females and older people [50] compared to younger people might have higher morbidity and mortality from COVID-19–related diseases, and they are more willing to be vaccinated [51].

Overall, only paying excess attention blindly cannot effectively allay the diverse public concerns and fears about vaccines. Specialized interventions should be implemented to address these concerns raised by different populations. Some studies in the field of public health are worthy of reference. For example, Brooke et al [52] discussed the problems encountered by older people during COVID-19 and put forward some practical suggestions for older people with comprehensive health help, both psychological and physical in COVID-19, the development of social network communications through online technologies (such as Facebook, Twitter, WhatsApp, and other similar platforms), and some creative ways of virtual entertainment (such as performances of symphony orchestras in virtual concert halls). Malik et al [51] pointed out that to build confidence in COVID-19 vaccines, thoughtful and targeted messaging and education need to be developed, not only for the general American population, but also specifically for high-risk groups. According to this, fantastic short videos or simple messages related to vaccines could be recommended to population groups with low sentiments toward vaccines to arouse their interest and educate them about the safety and effectiveness of vaccines. Additionally, we found that organizations, individuals with occupations of the 1st category, those with account age ≥10 years, and those with follower number ≥5000 were more concentrated and positive toward COVID-19 vaccines, as shown in Figure 10. Inspired by this, we further suggest that public health authorities cooperate with these users by using their social influence to expand the publicity of vaccines.

Limitations

Due to the high complexity of multilingual analysis and the insufficient support for detecting the various characteristics of the population groups, this paper only extracted data in English and key demographic characteristics for analysis. In addition, considering that too many OCs may reduce the accuracy of the occupation predictor, we only divided the occupations into 3 categories, which may have resulted in the loss of some fine-grained information. Despite these limitations, it did not affect the overall findings.

Conclusions

By analyzing large-scale tweets during vaccine development and vaccination, this study tracked the year-long evolution of attitudes toward COVID-19 vaccines among population groups, and offered rich evidence to gain insights about the attitude patterns of the international population on social media. Through well-organized approaches, governments, public health agencies, health care providers, and influential Twitter users can work together to help those populations with low sentiments get through this difficult period. At last, it is worth mentioning that the method applied in this paper can be easily extended to other public health events for multidimensional and large-scale research on the long-term evolution of human responses. The source code developed in this study is available for use at GitHub [18].

Acknowledgments

This research was supported by the National Natural Science Foundation of China (grants 61876150 and 12026609), the Key Research and Development Program of Shaanxi Province (grant number 2021SF-188), the Science and Technology Program of the City of Xi'an (grant number 20YXYJ0009-12), and the Youth Fund of the Second Affiliated Hospital of Xi'an Jiaotong University (grant number YJ(QN)202004).

Conflicts of Interest

None declared.

Multimedia Appendix 1
Supplemental information about vaccine-related news between March and April 2021.
[DOCX File , 20 KB - jmir_v24i1e32394_app1.docx]

Multimedia Appendix 2
Supplemental information about the sentiments toward vaccines at 5 major vaccine-related events.
[DOCX File , 21 KB - jmir_v24i1e32394_app2.docx]

References

1. What's the difference between quarantine and isolation? Mayo Clinic. URL: https://www.mayoclinic.org/diseases-conditions/coronavirus/in-depth/coronavirus-quarantine-and-isolation/art-20484503 [accessed 2021-04-20]

2. Randolph HE, Barreiro LB. Herd Immunity: Understanding COVID-19. Immunity 2020 May 19;52(5):737-741 [FREE Full text] [doi: 10.1016/j.immuni.2020.04.012] [Medline: 32433946]
3. Krammer F. SARS-CoV-2 vaccines in development. Nature 2020 Oct;586(7830):516-527. [doi: 10.1038/s41586-020-2798-3] [Medline: 32967006]
4. COVID-19 Global literature on coronavirus disease. World Health Organization. URL: https://search.bvsalud.org/global-literature-on-novel-coronavirus-2019-ncov/ [accessed 2021-04-01]
5. Lazarus JV, Ratzan SC, Palayew A, Gostin LO, Larson HJ, Rabin K, et al. A global survey of potential acceptance of a COVID-19 vaccine. Nat Med 2021 Feb;27(2):225-228 [FREE Full text] [doi: 10.1038/s41591-020-1124-9] [Medline: 33082575]
6. Khubchandani J, Sharma S, Price JH, Wiblishauser MJ, Sharma M, Webb FJ. COVID-19 Vaccination Hesitancy in the United States: A Rapid National Assessment. J Community Health 2021 Apr 03;46(2):270-277 [FREE Full text] [doi: 10.1007/s10900-020-00958-x] [Medline: 33389421]
7. Petravić L, Arh R, Gabrovec T, Jazbec L, Rupčić N, Starešinič N, et al. Factors Affecting Attitudes towards COVID-19 Vaccination: An Online Survey in Slovenia. Vaccines (Basel) 2021 Mar 12;9(3):247 [FREE Full text] [doi: 10.3390/vaccines9030247] [Medline: 33808958]
8. Bachtiger P, Adamson A, Chow J, Sisodia R, Quint JK, Peters NS. The Impact of the COVID-19 Pandemic on the Uptake of Influenza Vaccine: UK-Wide Observational Study. JMIR Public Health Surveill 2021 Apr 14;7(4):e26734 [FREE Full text] [doi: 10.2196/26734] [Medline: 33651708]
9. Chen H, Li X, Gao J, Liu X, Mao Y, Wang R, et al. Health Belief Model Perspective on the Control of COVID-19 Vaccine Hesitancy and the Promotion of Vaccination in China: Web-Based Cross-sectional Study. J Med Internet Res 2021 Sep 06;23(9):e29329 [FREE Full text] [doi: 10.2196/29329] [Medline: 34280115]
10. Hou Z, Tong Y, Du F, Lu L, Zhao S, Yu K, et al. Assessing COVID-19 Vaccine Hesitancy, Confidence, and Public Engagement: A Global Social Listening Study. J Med Internet Res 2021 Jun 11;23(6):e27632 [FREE Full text] [doi: 10.2196/27632] [Medline: 34061757]
11. Lee H, Noh EB, Park SJ, Nam HK, Lee TH, Lee GR, et al. COVID-19 Vaccine Perception in South Korea: Web Crawling Approach. JMIR Public Health Surveill 2021 Sep 08;7(9):e31409 [FREE Full text] [doi: 10.2196/31409] [Medline: 34348890]
12. Pullan S, Dey M. Vaccine hesitancy and anti-vaccination in the time of COVID-19: A Google Trends analysis. Vaccine 2021 Apr 01;39(14):1877-1881 [FREE Full text] [doi: 10.1016/j.vaccine.2021.03.019] [Medline: 33715904]
13. An L, Russell DM, Mihalcea R, Bacon E, Huffman S, Resnicow K. Online Search Behavior Related to COVID-19 Vaccines: Infodemiology Study. JMIR Infodemiology 2021 Nov 12;1(1):e32127 [FREE Full text] [doi: 10.2196/32127] [Medline: 34841200]
14. Yin F, Wu Z, Xia X, Ji M, Wang Y, Hu Z. Unfolding the Determinants of COVID-19 Vaccine Acceptance in China. J Med Internet Res 2021 Jan 15;23(1):e26089 [FREE Full text] [doi: 10.2196/26089] [Medline: 33400682]
15. Lyu JC, Han EL, Luli GK. COVID-19 Vaccine-Related Discussion on Twitter: Topic Modeling and Sentiment Analysis. J Med Internet Res 2021 Jun 29;23(6):e24435 [FREE Full text] [doi: 10.2196/24435] [Medline: 34115608]
16. Praveen S, Ittamalla R, Deepak G. Analyzing the attitude of Indian citizens towards COVID-19 vaccine - A text analytics study. Diabetes Metab Syndr 2021 Mar;15(2):595-599 [FREE Full text] [doi: 10.1016/j.dsx.2021.02.031] [Medline: 33714134]
17. Liu S, Liu J. Public attitudes toward COVID-19 vaccines on English-language Twitter: A sentiment analysis. Vaccine 2021 Sep 15;39(39):5499-5505 [FREE Full text] [doi: 10.1016/j.vaccine.2021.08.058] [Medline: 34452774]
18. Attitudes of COVID-19 Vaccines. GitHub. URL: https://github.com/cyzhang87/AttitudesOfCOVID-19Vaccines [accessed 2021-07-14]
19. Volume streams. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/twitter-api/tweets/sampled-stream/introduction [accessed 2021-07-01]
20. Chen E, Lerman K, Ferrara E. Tracking Social Media Discourse About the COVID-19 Pandemic: Development of a Public Coronavirus Twitter Data Set. JMIR Public Health Surveill 2020 May 29;6(2):e19273 [FREE Full text] [doi: 10.2196/19273] [Medline: 32427106]
21. Developer Agreement and Policy. Twitter Developer Platform. URL: https://developer.twitter.com/en/developer-terms/agreement-and-policy [accessed 2021-07-01]
22. Tweets lookup. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/twitter-api/tweets/lookup/api-reference/get-tweets [accessed 2021-07-01]
23. Users lookup. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/twitter-api/users/lookup/introduction [accessed 2021-07-01]
24. COVID-19 stream. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/labs/covid19-stream/filtering-rules [accessed 2021-07-01]
25. Wang Z, Hale S, Adelani D, Grabowicz P, Hartman T, Flöck F, et al. Demographic Inference and Representative Population Estimates from Multilingual Social Media Data. In: WWW '19: The World Wide Web Conference. 2019 Presented at: The World Wide Web Conference; May 13-17, 2019; San Francisco, CA, USA p. 2056-2067. [doi: 10.1145/3308558.3313684]

26. Rothe R, Timofte R, Van Gool L. Deep Expectation of Real and Apparent Age from a Single Image Without Facial Landmarks. Int J Comput Vis 2016 Aug 10;126(2-4):144-157. [doi: 10.1007/s11263-016-0940-3]
27. Zhang C, Xu S, Li Z, Hu S. Understanding Concerns, Sentiments, and Disparities Among Population Groups During the COVID-19 Pandemic Via Twitter Data Mining: Large-scale Cross-sectional Study. J Med Internet Res 2021 Mar 05;23(3):e26482 [FREE Full text] [doi: 10.2196/26482] [Medline: 33617460]
28. Preoţiuc-Pietro D, Lampos V, Aletras N. An analysis of the user occupational class through Twitter content. In: Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 2015 Presented at: 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing; July 2015; Beijing, China. [doi: 10.3115/v1/p15-1169]
29. Standard Occupational Classification. Office for National Statistics. URL: https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc [accessed 2021-04-01]
30. Pan J, Bhardwaj R, Lu W, Chieu HL, Pan X, Puay NY. Twitter Homophily: Network Based Prediction of User's Occupation. In: Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 2019 Presented at: 57th Annual Meeting of the Association for Computational Linguistics; July 2019; Florence, Italy. [doi: 10.18653/v1/p19-1252]
31. Lyu H, Wang J, Wu W, Duong V, Zhang X, Dye TD, et al. Social media study of public opinions on potential COVID-19 vaccines: informing dissent, disparities, and dissemination. Intell Med 2021 Aug 25 [FREE Full text] [doi: 10.1016/j.imed.2021.08.001] [Medline: 34457371]
32. Get information about a place. Twitter Developer Platform. URL: https://developer.twitter.com/en/docs/twitter-api/v1/geo/place-information/api-reference/get-geo-id-place_id [accessed 2021-04-01]
33. GeoPy. URL: https://geopy.readthedocs.io/en/stable/ [accessed 2021-04-01]
34. pycountry. pypi. URL: https://pypi.org/project/pycountry/ [accessed 2021-04-01]
35. Hutto C, Gilbert E. Vader: A parsimonious rule-based model for sentiment analysis of social media text. comp.social lab. URL: http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf [accessed 2022-01-12]
36. Colneric N, Demsar J. Emotion Recognition on Twitter: Comparative Study and Training a Unison Model. IEEE Trans. Affective Comput 2020 Jul 1;11(3):433-446. [doi: 10.1109/taffc.2018.2807817]
37. Ekman P. An argument for basic emotions. Cognition and Emotion 2008 Jan 07;6(3-4):169-200. [doi: 10.1080/02699939208411068]
38. Wuensch KL, Evans JD. Straightforward Statistics for the Behavioral Sciences. Journal of the American Statistical Association 1996 Dec;91(436):1750. [doi: 10.2307/2291607]
39. Coronavirus Pandemic (COVID-19). Our World In Data. URL: https://ourworldindata.org/coronavirus [accessed 2021-04-01]
40. Jackson LA, Anderson EJ, Rouphael NG, Roberts PC, Makhene M, Coler RN, et al. An mRNA Vaccine against SARS-CoV-2 — Preliminary Report. N Engl J Med 2020 Nov 12;383(20):1920-1931. [doi: 10.1056/nejmoa2022483]
41. Mulligan MJ, Lyke KE, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Phase I/II study of COVID-19 RNA vaccine BNT162b1 in adults. Nature 2020 Oct 12;586(7830):589-593. [doi: 10.1038/s41586-020-2639-4] [Medline: 32785213]
42. Gallagher J. Covid vaccine: First 'milestone' vaccine offers 90% protection. BBC NEWS. URL: https://www.bbc.com/news/health-54873105 [accessed 2021-04-01]
43. Coronavirus disease (COVID-19): Vaccines. World Health Organization. URL: https://www.who.int/news-room/q-a-detail/coronavirus-disease-(covid-19)-vaccines [accessed 2021-04-01]
44. Hu D, Lou X, Xu Z, Meng N, Xie Q, Zhang M, et al. More effective strategies are required to strengthen public awareness of COVID-19: Evidence from Google Trends. J Glob Health 2020 Jun;10(1):011003 [FREE Full text] [doi: 10.7189/jogh.10.011003] [Medline: 32373339]
45. Yan C, Law M, Nguyen S, Cheung J, Kong J. Comparing Public Sentiment Toward COVID-19 Vaccines Across Canadian Cities: Analysis of Comments on Reddit. J Med Internet Res 2021 Sep 24;23(9):e32685 [FREE Full text] [doi: 10.2196/32685] [Medline: 34519654]
46. Hu T, Wang S, Luo W, Zhang M, Huang X, Yan Y, et al. Revealing public opinion towards COVID-19 vaccines using Twitter data in the United States: a spatiotemporal perspective. medRxiv. URL: https://www.medrxiv.org/content/10.1101/2021.06.02.21258233v1 [accessed 2022-01-12]
47. 5 Things To Know About the Delta Variant. YaleMedicine. URL: https://www.yalemedicine.org/news/5-things-to-know-delta-variant-covid [accessed 2021-10-20]
48. Paul E, Steptoe A, Fancourt D. Attitudes towards vaccines and intention to vaccinate against COVID-19: Implications for public health communications. Lancet Reg Health Eur 2021 Feb;1:100012 [FREE Full text] [doi: 10.1016/j.lanepe.2020.100012] [Medline: 33954296]
49. Peckham H, de Gruijter NM, Raine C, Radziszewska A, Ciurtin C, Wedderburn LR, et al. Male sex identified by global COVID-19 meta-analysis as a risk factor for death and ITU admission. Nat Commun 2020 Dec 09;11(1):6317 [FREE Full text] [doi: 10.1038/s41467-020-19741-6] [Medline: 33298944]
50. Mueller AL, McNamara MS, Sinclair DA. Why does COVID-19 disproportionately affect older people? Aging (Albany NY) 2020 May 29;12(10):9959-9981 [FREE Full text] [doi: 10.18632/aging.103344] [Medline: 32470948]

51. Malik AA, McFadden SM, Elharake J, Omer SB. Determinants of COVID-19 vaccine acceptance in the US. EClinicalMedicine 2020 Sep;26:100495 [FREE Full text] [doi: 10.1016/j.eclinm.2020.100495] [Medline: 32838242]
52. Brooke J, Jackson D. Older people and COVID-19: Isolation, risk and ageism. J Clin Nurs 2020 Jul 02;29(13-14):2044-2046. [doi: 10.1111/jocn.15274] [Medline: 32239784]

Abbreviations

API: application programming interface
M3: multimodal, multilingual, and multiattribute
OC: occupation category
OR: odds ratio
SOC: Standard Occupation Classification
VADER: Valence Aware Dictionary and Emotional Reasoner

Edited by C Basch; submitted 28.07.21; peer-reviewed by Z Hu, C Giraud-Carrier; comments to author 20.09.21; revised version received 04.11.21; accepted 03.12.21; published 21.01.22.

Please cite as:
Zhang C, Xu S, Li Z, Liu G, Dai D, Dong C
The Evolution and Disparities of Online Attitudes Toward COVID-19 Vaccines: Year-long Longitudinal and Cross-sectional Study
J Med Internet Res 2022;24(1):e32394
URL: https://www.jmir.org/2022/1/e32394
doi:10.2196/32394
PMID:34878410

Original Paper

Patterns of Symptom Tracking by Caregivers and Patients With Dementia and Mild Cognitive Impairment: Cross-sectional Study

Taylor Dunn[1], MSc; Susan E Howlett[1,2,3], PhD; Sanja Stanojevic[1,4], PhD; Aaqib Shehzad[1], BDS, MHI; Justin Stanley[1], BEng; Kenneth Rockwood[1,3,5], MD

[1]Ardea Outcomes, Halifax, NS, Canada

[2]Department of Pharmacology, Dalhousie University, Halifax, NS, Canada

[3]Division of Geriatric Medicine, Dalhousie University, Halifax, NS, Canada

[4]Department of Community Health and Epidemiology, Dalhousie University, Halifax, NS, Canada

[5]Geriatric Medicine Research Unit, Nova Scotia Health Authority, Halifax, NS, Canada

Corresponding Author:
Kenneth Rockwood, MD
Geriatric Medicine Research Unit
Nova Scotia Health Authority
1421-5955 Veterans Memorial Lane
Halifax, NS, B3H 2E1
Canada
Phone: 1 902 473 8687
Email: research@ardeaoutcomes.com

Abstract

Background: Individuals with dementia and mild cognitive impairment (MCI) experience a wide variety of symptoms and challenges that trouble them. To address this heterogeneity, numerous standardized tests are used for diagnosis and prognosis. myGoalNav Dementia is a web-based tool that allows individuals with impairments and their caregivers to identify and track outcomes of greatest importance to them, which may be a less arbitrary and more sensitive way of capturing meaningful change.

Objective: We aim to explore the most frequent and important symptoms and challenges reported by caregivers and people with dementia and MCI and how this varies according to disease severity.

Methods: This cross-sectional study involved 3909 web-based myGoalNav users (mostly caregivers of people with dementia or MCI) who completed symptom profiles between 2006 and 2019. To make a symptom profile, users selected their most personally meaningful or troublesome dementia-related symptoms to track over time. Users were also asked to rank their chosen symptoms from least to most important, which we called the symptom potency. As the stage of disease for these web-based users was unknown, we applied a supervised staging algorithm, previously trained on clinician-derived data, to classify each profile into 1 of 4 stages: MCI and mild, moderate, and severe dementia. Across these stages, we compared symptom tracking frequency, symptom potency, and the relationship between frequency and potency.

Results: Applying the staging algorithm to the 3909 user profiles resulted in 917 (23.46%) MCI, 1596 (40.83%) mild dementia, 514 (13.15%) moderate dementia, and 882 (22.56%) severe dementia profiles. We found that the most frequent symptoms in MCI and mild dementia profiles were similar and comprised early hallmarks of dementia (eg, recent memory and language difficulty). As the stage increased to moderate and severe, the most frequent symptoms were characteristic of loss of independent function (eg, incontinence) and behavioral problems (eg, aggression). The most potent symptoms were similar between stages and generally reflected disruptions in everyday life (eg, problems with hobbies or games, travel, and looking after grandchildren). Symptom frequency was negatively correlated with potency at all stages, and the strength of this relationship increased with increasing disease severity.

Conclusions: Our results emphasize the importance of patient-centricity in MCI and dementia studies and illustrate the valuable real-world evidence that can be collected with digital tools. Here, the most frequent symptoms across the stages reflected our understanding of the typical disease progression. However, the symptoms that were ranked as most personally important by users were generally among the least frequently selected. Through individualization, patient-centered instruments such as myGoalNav can complement standardized measures by capturing these infrequent but potent outcomes.

(J Med Internet Res 2022;24(1):e29219) doi:10.2196/29219

KEYWORDS

dementia; mild cognitive impairment; real-world evidence; patient-centric outcomes; machine learning; dementia stage; Alzheimer disease; symptom tracking

Introduction

Background

It is proving difficult to understand what constitutes successfully treated late-life dementia. This reflects in part the evolving understanding of Alzheimer disease (AD). Contemporary thinking sees AD as a biological construct, defined by biomarkers that can be detected in vivo or during autopsy [1]. In this formulation, AD is distinguished from Alzheimer dementia, a clinical syndrome [2]. However, this separation is not clear. In contrast to the prior view that a definitive diagnosis of AD could only be made at autopsy [3], it is now recognized that many people who meet the neuropathological criteria for AD do not have dementia when alive [4]. Few people with late-life dementia have *pure* AD; in the great majority, it is present along with many other neuropathological features [5,6]. Furthermore, even a full suite of neuropathological markers cannot distinguish between people who had dementia when alive and those who did not; other factors such as history of delirium [7], prior hospitalization [8], and degree of frailty are each important [9]. Similarly, a range of factors, from the level of education [10] to stimulating psychosocial and lifestyle experiences [11], is seen as potentially protective, even if less well-studied. A further challenge to defining successful treatment is that standard outcome measures, notably including the commonly used AD Assessment Scale–Cognitive subscale (ADAS-Cog), can underestimate meaningful clinical changes [12,13].

The new consensus on defining AD and the broader understanding of what gives rise to late-life dementia together have propelled a rethinking of which outcomes to measure in dementia and predementia clinical trials [14,15]. The Food and Drug Administration guidelines in 2018 [16] suggested that a single primary end point, which assesses both cognitive and functional effects (eg, the Clinical Dementia Rating [CDR] scale–Sum of Boxes [17,18]) may be used to evaluate treatment in early-stage patients with biomarker-defined AD. With this reevaluation, it may be useful to consider patient-reported impacts of treatment. This could be a less arbitrary means of understanding treatment efficacy compared with changes in biomarkers that tend to correlate poorly with clinical measures [19-24]. Along these lines, the lack of correlation between clinical manifestations of the disease and biomarker positivity has motivated the reconsideration of a purely biological definition of AD, suggesting that the disease designation be restricted to people who combine biomarker positivity with specific AD phenotypes [25].

Patient-centric outcome measures, in which patient (and caregiver) preferences are directly incorporated and measured, have slowly gained traction in clinical trials and research communities [26], including an endorsement from the Food and Drug Administration [27]. By giving a voice to the patient, we can achieve more meaningful and interpretable measures of treatment benefit [28], as seen in some dementia research, including clinical trials of people with AD receiving donepezil [29] and galantamine [30], which used goal attainment scaling [31,32] as a primary outcome. Here, personalized outcomes offered highly sensitive measures of change that were viewed as clinically meaningful by patients and caregivers [13]. This approach also provided additional insights into what is most important to this population, for example, an unanticipated treatment benefit in the troubling symptom of verbal repetition [33,34]. Indeed, from mild cognitive impairment (MCI), which is the symptomatic predementia stage of AD [35], to the severe stage of dementia, patients are troubled by diverse sets of cognitive, functional, and behavioral symptoms. However, we lack a comprehensive inventory of symptoms across the disease spectrum and their susceptibility to treatment. Surveys of symptoms are few, in part as they are expensive. For this purpose, the web-based environment can be well-suited.

Our group has shown that data on people living with MCI and dementia and their caregivers can be acquired with an internet-based tool called myGoalNav Dementia (previously SymptomGuide Dementia; developed by Ardea Outcomes). This symptom tracking platform provides a large library from which users can identify and track dementia symptoms that are most important to them [36]. Note the distinction between a symptom being present—as in a *tick box* survey of symptom prevalence—and one being important to individual patients and caregivers. Earlier, we had used myGoalNav to investigate construct validity with the Dependence Scale [37]; identify clusters of neuropsychiatric symptoms [38]; characterize the symptoms of verbal repetition [39], misplacing objects [40], and agitation [41]; and evaluate donepezil in a 6-month open-label study [42]. Here, we use myGoalNav Dementia to better understand the patterns of dementia symptom tracking with the severity of impairment, as staged by a machine learning algorithm.

Objective

The aims of this cross-sectional study are three-fold: (1) to compare symptom frequency by stage, (2) to compare symptom importance by stage, and (3) to examine the relationship between frequency and importance. In doing so, our overall goal is to demonstrate the usefulness and the types of insights gained from web-based symptom tracking in people with dementia and MCI.

Methods

Data Collection

The data are from the myGoalNav Dementia platform, previously called SymptomGuide Dementia. Launched as a website in 2007 for people with cognitive impairment and their caregivers, the key feature of the platform is a library (or *menu*) of common dementia-related symptoms and challenges. The

library was developed over many years, beginning with a qualitative analysis of personalized treatment goals set by patients, caregivers, and clinicians in 2 clinical trials of anticholinesterase inhibitors [29,30] and from a memory clinic in Halifax, Nova Scotia. From this qualitative analysis, an expert geriatrician panel reviewed the first draft using the Delphi method [43] and arrived at a library of 60 symptoms [37,41]. In 2018, myGoalNav Dementia was redesigned as an iOS and Android mobile app, and based largely on user feedback, the library was expanded to 67 symptoms, each with 2 to 12 (median 9) plain language descriptors that provide an additional level of detail into symptom manifestation.

In addition to providing users with educational information and management tips for each symptom, users have the option of choosing from the library any number of symptoms and relevant descriptors to track over time, which are important to them or the person for whom they care. If they wish for further personalization, users can log *other* symptoms and descriptors that do not appear in the library. The set of initial symptoms selected by a user is called their baseline symptom profile, which they may supplement with additional demographic information such as age, gender, and living arrangements. As an optional step, users are also asked to rank their chosen symptoms from most to least important or troublesome.

myGoalNav Dementia is currently being retooled as a mobile-first web-based care app that better facilitates shared decision-making and improves the quantity and quality of touchpoints between the provider and patient. The transition to web-based technologies affords us flexibility in incorporating our dementia staging model within the app in the future. Currently, it is slated to undergo pilot testing with the collaboration of our health care provider partners and is no longer available as a community app.

myGoalNav was not designed to be an inventory of every dementia-related problem that an individual might experience. Rather, the library facilitates the selection of those symptoms that are most meaningful to each participant. For this analysis, we excluded outlier profiles created by individuals who chose >22 symptoms (95th percentile).

Staging Dementia

In 2013, we developed an artificial neural network model to stage dementia, which was trained on data from 320 memory clinic patients [42]. That model was updated in 2020 using a support vector machine–supervised learning algorithm trained on 717 patients [44]. Data from these patients were captured with myGoalNav in a memory clinic, a long-term care study [45], and a dementia clinical trial [42]. Patients were staged by a clinician using either the Functional Assessment Staging Test or the Global Deterioration Scale into 1 of 4 stages: MCI or mild, moderate, or severe dementia. Patient age and their symptom profiles served as inputs to train the model to predict the dementia stage. Further details of the model, including algorithm choice and performance, can be found in the study by Shehzad et al [44].

Statistical Analysis

User characteristics and demographics were summarized, and the differences between stages were tested. Categorical variables were summarized as percentages of users and tested using the Pearson chi-square test. Continuous variables were summarized as means and SDs or medians (lower and upper quartiles) and tested using the Kruskal–Wallis *H* test.

To compare symptom frequency by disease stage (objective 1), we first fit a logistic regression model with the number of profiles selecting each symptom in each stage as the dependent variable. Symptom name, stage, and the interaction between the 2 were included as independent variables. The estimates from this model were transformed to stage-specific symptom frequencies with 95% CIs.

We investigated the frequency differences between the stages in 2 ways. First, we computed Pearson correlation coefficient *r* on frequencies between each pair of stages, where a higher *r* coefficient indicates greater similarity in symptom selection. Second, we quantified the degree to which a symptom was associated with increasing or decreasing disease severity. This was accomplished by modifying our logistic regression model so that the stage is treated as a monotonic predictor variable [46] rather than a categorical variable without ordering. The estimates from this model can be interpreted as the average difference in frequency (on the log-odds scale) between adjacent stages (*Supplementary Methods* section in Multimedia Appendix 1 [46-49]).

To compare symptom importance by stage (objective 2), we began by defining *relative* symptom importance within a symptom profile as the weighted rank or *potency*:

$$w_{ij}=r_{ij}/n_j \quad \textbf{(1)}$$

where n_j is the number of symptoms in profile j, and r_{ij} is the rank (out of n_j) given to symptom I. Higher potency w_{ij} corresponds to the higher relative importance of symptom i to the user of profile j.

Next, to model this proportion while accounting for the wide range in the number of tracked symptoms among myGoalNav users, we used logistic regression with w_{ij} as the dependent variable; n_j as case weights; and categorical independent variables of symptom name, stage, and their interaction. As with the frequency analysis, we estimated pairwise similarity in stages by computing Pearson correlation coefficients between potency estimates.

To investigate the relationship between symptom frequency and potency (objective 3), we visualized the relationship by plotting frequency against potency estimates. We quantified the strength of these relationships using Pearson correlation.

No missing data were imputed for this study. As mentioned, we removed outlier profiles with >22 symptoms (95th percentile) but otherwise did not exclude users for missing or abnormal data. All analyses were performed in R version 4.0.2 (R Foundation for Statistical Computing) [50] using tidyverse packages [51].

Ethics

Clinic data were collected after having obtained written informed consent. Participants completed a form that allowed their anonymized data to be analyzed for research purposes. Data collection was approved by the research ethics committee of the Nova Scotia Health Authority. myGoalNav users consented to terms of use, which included allowing their data to be aggregated and used for research purposes. Users were assured that the research findings would be presented in a manner that would not disclose personally identifying information.

Data Availability

Aggregated data are presented in Table S1 (symptom frequency) and Table S2 (symptom potency) of Multimedia Appendix 1. For confidentiality reasons, user-level data cannot be made publicly available. Access to deidentified data may be provided upon reasonable request.

Results

The Sample

To date, 12,347 users have signed up for myGoalNav, and of these, 4213 (34.12%) users created a symptom profile. Of the 4213 users, after removing the profiles tracking >22 symptoms, our final sample size was 3909 (92.78%) profiles, with creation dates ranging from May 15, 2006, to November 15, 2018.

The great majority, 96.01% (3753/3909), of these symptom profiles were made on the web platform, with 3.99% (156/3909) from the later mobile app. Most profiles (3792/3909, 97.01%) were completed by caregivers, and the remaining (117/3909, 2.99%) were completed by participants (people with cognitive impairment) on their own behalf. The staging algorithm led to the following distribution of severity across the 3909 profiles: 917 (23.46%) MCI, 1596 (40.82%) mild dementia, 514 (13.15%) moderate dementia, and 882 (22.56%) severe dementia.

Participant characteristics and demographics are summarized by stage in Table 1. With increasing severity from MCI to moderate dementia, participants tended to be older, less well-educated, more likely to identify as women, and less likely to live on their own. A minority of caregivers (1361/3792, 35.89% users) also provided information about themselves. Most caregivers were women (515/802, 64.2%), aged 46-55 years (184/583, 31.6%), and spouses or partners of the participant (156/524, 29.8%).

Table 1. Baseline characteristics of the myGoalNav participants, stratified by stage (N=3909).

Characteristic[a]	Total	MCI[b] (n=917)	Mild (n=1596)	Moderate (n=514)	Severe (n=882)	Test statistic[c]		
						P value	*H* test (*df*)	Chi-square (*df*)
Age (years; n=2473), mean (SD)	75.4 (12.4)	70.5 (13.2)	74.5 (12.1)	80.9 (8.7)	78.2 (11.9)	<.001	192.8 (3)	N/A[d]
Gender (n=3909), n (%)						.046	N/A	8.0 (3)
Man	976 (37.7)[e]	228 (40.9)[f]	431 (39)[g]	127 (33.7)[h]	190 (34.7)[i]			
Woman	1611 (62.3)[e]	329 (59.1)[f]	674 (61)[g]	250 (66.3)[h]	358 (65.3)[i]			
Number of symptoms (n=2587), median (Q1-Q3)	4 (2-7)	2 (1-4)	5 (3-8)	7 (4-11)	4 (2-7)	<.001	669.7 (3)	N/A
Education (n=1337), n (%)						.18	N/A	12.7 (9)
Secondary school or less	625 (46.7)	117 (40.8)[j]	283 (45.7)[k]	99 (52.1)[l]	126 (52.3)[m]			
Trade school	71 (5.3)	14 (4.9)[j]	32 (5.2)[k]	9 (4.7)[l]	16 (6.6)[m]			
Undergraduate	439 (32.8)	108 (37.6)[j]	211 (34.1)[k]	53 (27.9)[l]	67 (27.8)[m]			
Graduate	202 (15.1)	48 (16.7)[j]	93 (15)[k]	29 (15.3)[l]	32 (13.3)[m]			
Living arrangement (n=2013), n (%)						<.001	N/A	126.5 (12)
Alone	290 (14.4)	67 (14.9)[n]	143 (15.8)[o]	35 (12.2)[p]	45 (12.1)[q]			
Assisted living	315 (15.6)	39 (8.7)[n]	102 (11.3)[o]	71 (24.8)[p]	103 (27.8)[q]			
With caregiver	335 (16.6)	59 (13.1)[n]	153 (16.9)[o]	54 (18.9)[p]	69 (18.6)[q]			
With family or friend	1038 (51.6)	283 (62.9)[n]	494 (54.5)[o]	120 (42)[p]	141 (38)[q]			
With paid companion	35 (1.7)	2 (0.4)[n]	14 (1.5)[o]	6 (2.1)[p]	13 (3.5)[q]			

[a]N is the number of users with nonmissing values.

[b]MCI: mild cognitive impairment.

[c]Comparisons between stages: Pearson chi-square test and Kruskal–Wallis *H* test.

[d]N/A: not applicable.

[e]Sample size, n=2587.

[f]Sample size, n=557.

[g]Sample size, n=1105.

[h]Sample size, n=377.

[i]Sample size, n=548.

[j]Sample size, n=287.

[k]Sample size, n=619.

[l]Sample size, n=190.

[m]Sample size, n=241.

[n]Sample size, n=450.

[o]Sample size, n=906.

[p]Sample size, n=286.

[q]Sample size, n=371.

Symptom Frequency

Figure 1 depicts the 10 most frequent symptoms in each stage. In MCI, mild dementia, and moderate dementia profiles, the most common symptom was memory of recent events: 33.41% (1306/3909), 36.71% (1435/3909), and 36.99% (1446/3909) of profiles, respectively. This early hallmark of dementia was tracked much less often at the severe stage, where 9.79% (383/3909) of profiles showed it being tracked. Other symptoms were tracked more often with greater severity. For example, sleep disturbance tracking increased from 6.7% (61/917) in MCI profiles to 15.47% (247/1596) in mild dementia profiles and to 25.5% (131/514) in moderate dementia profiles and was the most frequently tracked symptom in the severe dementia profiles (215/882, 24.4%).

In addition to memory of recent events, MCI profiles were best characterized by repetitive questions or stories; no other

symptom had a tracking frequency >20%. Those with mild and moderate dementia profiles showed more variety in symptom selection, with 6 and 8 symptoms >20% frequency, whereas those with severe dementia profiles were slightly more uniform, with 4 symptoms above that mark.

How frequencies varied is also illustrated by changes in the correlations between pairs of symptoms drawn from adjacent stages (Figure 2). Using this metric, the most similar stages (ie, highest correlation coefficient) were MCI and mild dementia. Indeed, those with MCI and mild dementia profiles shared 80% (8/10) of their most frequent symptoms. To a lesser degree, individuals with moderate dementia profiles had symptom frequencies similar to MCI and mild dementia profiles. There were four symptoms shared among the top 10 of these three stages: attention or concentration, irritability or frustration, memory of recent events, and repetitive questions or stories. By a large margin, severe dementia profiles had the most distinct set of frequent symptoms, although the correlation increased with increasing severity: r=0.14, 0.25, and 0.48 in MCI, mild dementia, and moderate dementia profiles, respectively. The most frequent symptoms in severe dementia profiles were characteristic of loss of independent function (incontinence, mobility, eating, and personal care or hygiene) and more extreme behavioral problems (aggression, low mood, and delusions and paranoia).

Model-estimated monotonicity of frequency with dementia severity is shown for each symptom in Figure 3. More symptoms exhibited positive monotonicity (36 symptoms with lower 95% CI >0) than negative (7 symptoms with upper 95% CI <0). The 6 symptoms with the highest positive and negative monotonicity are shown on the right in Figure 3.

Overall frequency and frequency by stage for each symptom are summarized in Table S1 in Multimedia Appendix 1.

Figure 1. The 10 most frequent symptoms tracked by baseline myGoalNav profiles, stratified by stage. Data are presented as point estimates and 95% CIs from the logistic regression model. MCI: mild cognitive impairment.

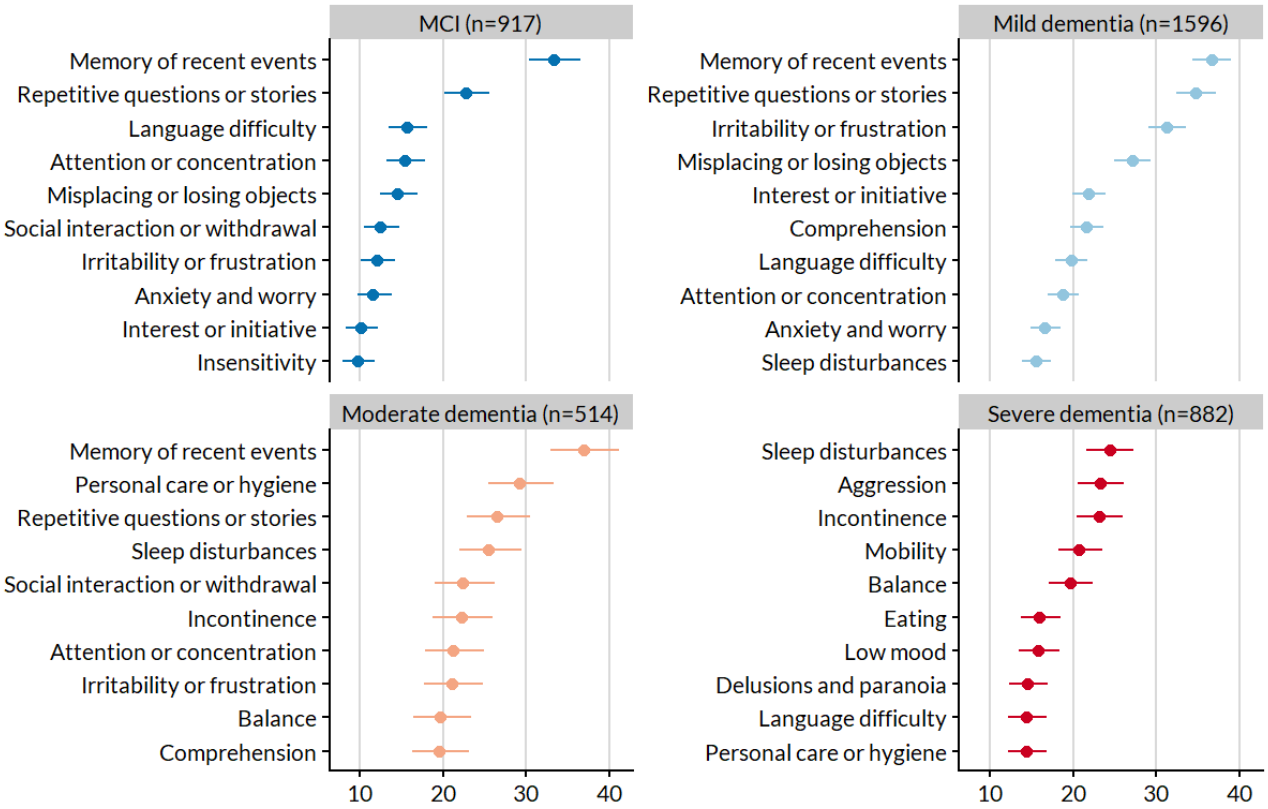


Figure 2. Pairwise relationships of symptom frequency between stages. Pearson correlation coefficients (*r*) and lines of best fit (with 95% CIs) are displayed for each pair. MCI: mild cognitive impairment.

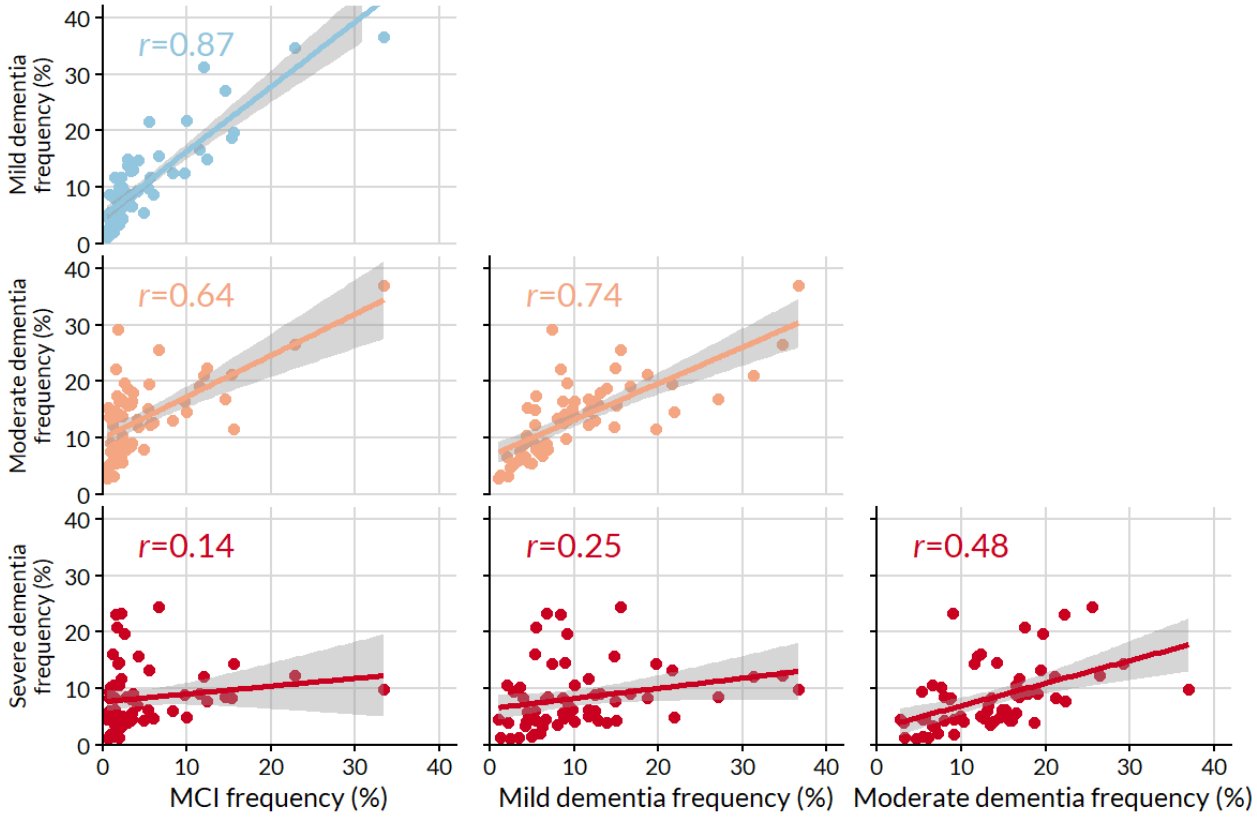
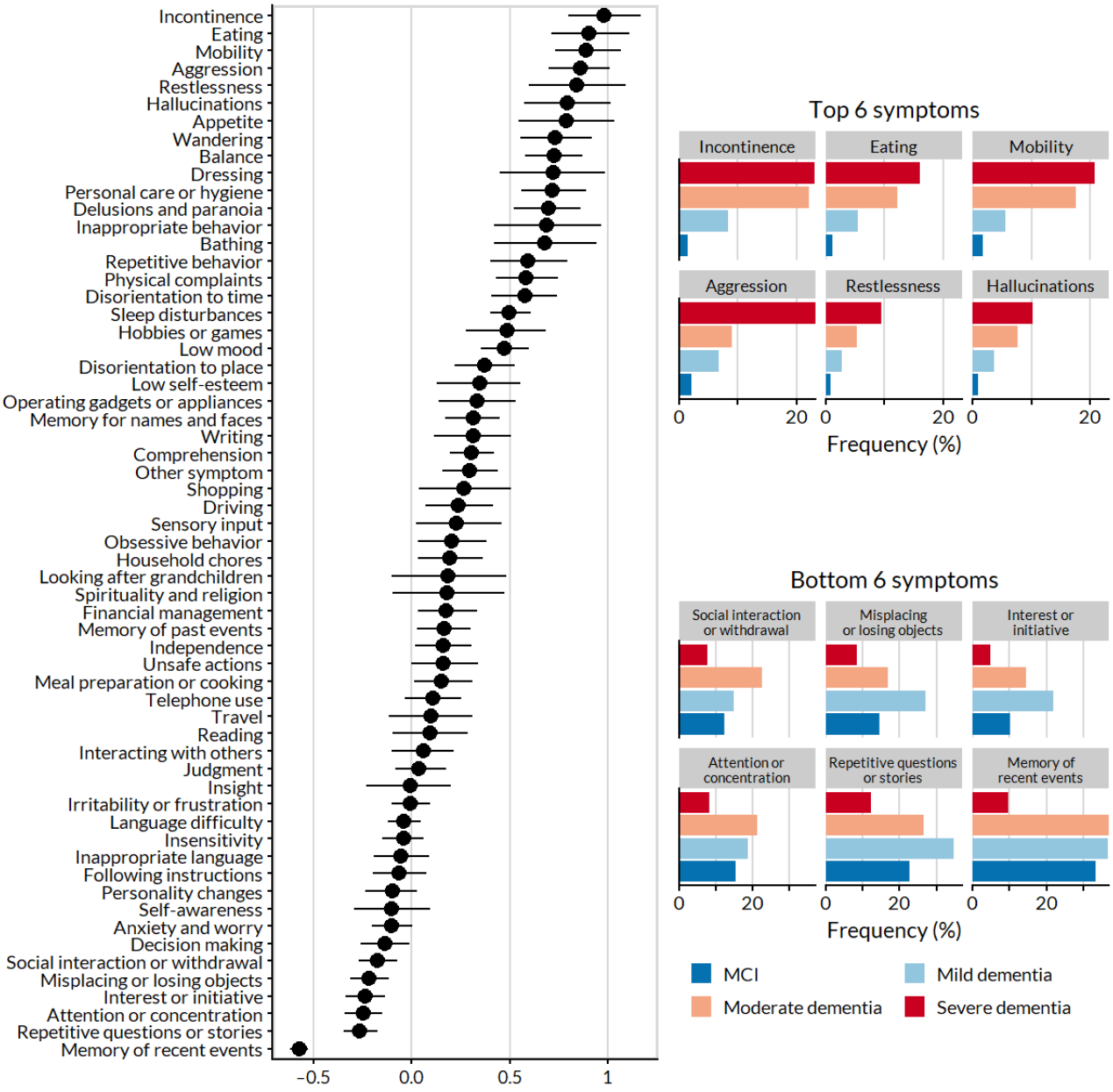


Figure 3. Estimated symptom monotonicity, where higher values indicate the increasing frequency with ordered stage. Data are presented as point estimates and 95% CIs from the logistic regression model, with stage as a monotonic predictor (left). Stage-specific frequencies for the 6 symptoms with the highest positive monotonicity and the 6 symptoms with the highest negative monotonicity (right). MCI: mild cognitive impairment.



Symptom Potency

As the ranking of symptoms is not compulsory on myGoalNav, the potency analysis involved 2874 symptom profiles (632, 21.99% MCI, 1207, 41.99% mild dementia, 402, 13.98% moderate dementia, and 632, 21.99% severe dementia). The model estimates of the 10 most potent symptoms by stage are shown in Figure 4, and the rest can be found in Table S2 in Multimedia Appendix 1. Of the 2874 symptoms, 2 (0.07%) symptoms stood out as important regardless of stage—travel and looking after grandchildren—which were among the top 3 most potent symptoms in each stage.

Figure 5 shows the pairwise relationship of symptom potency between stages. The most similar pairs of stages were moderate dementia profiles with mild and severe dementia profiles. The greatest differences in potency were MCI profiles with moderate and severe dementia profiles.

Figure 4. The 10 most potent symptoms tracked by baseline myGoalNav profiles, stratified by stage. Data are presented as point estimates and 95% CIs from the logistic regression model. MCI: mild cognitive impairment.

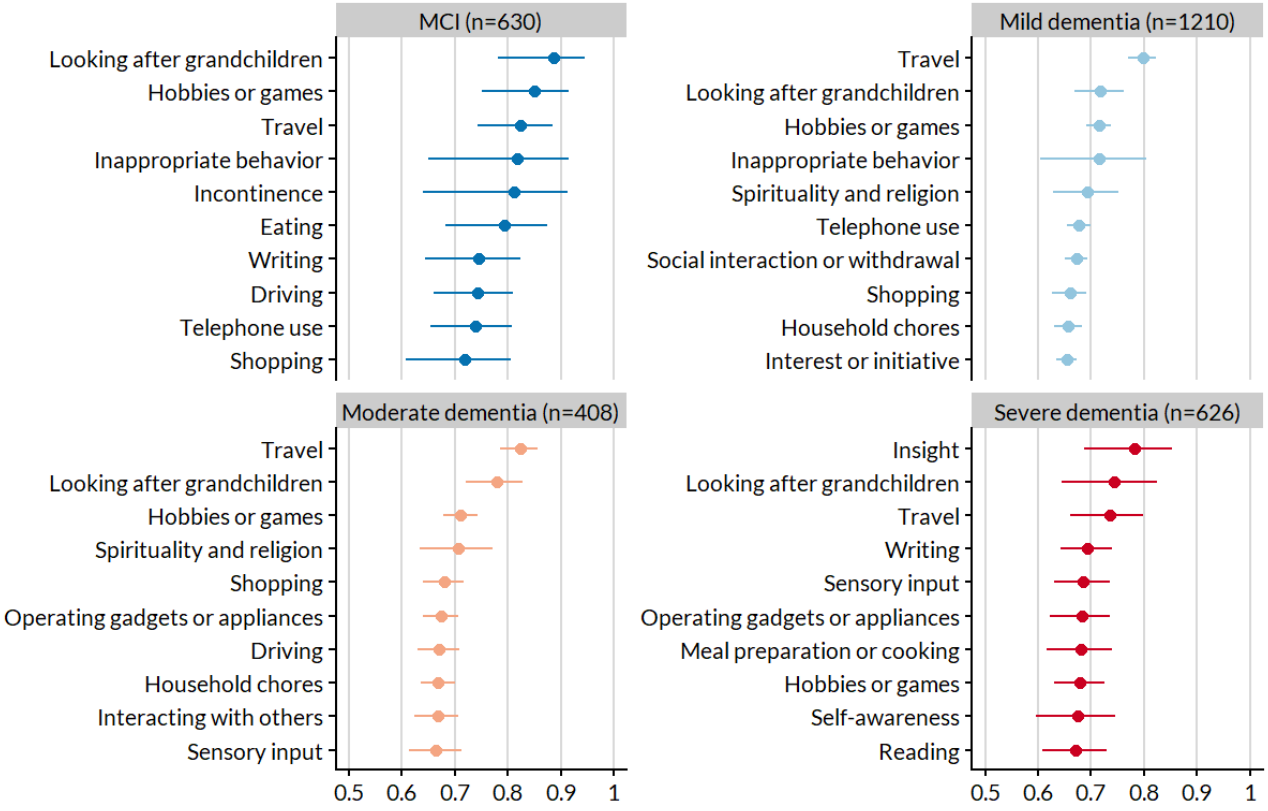
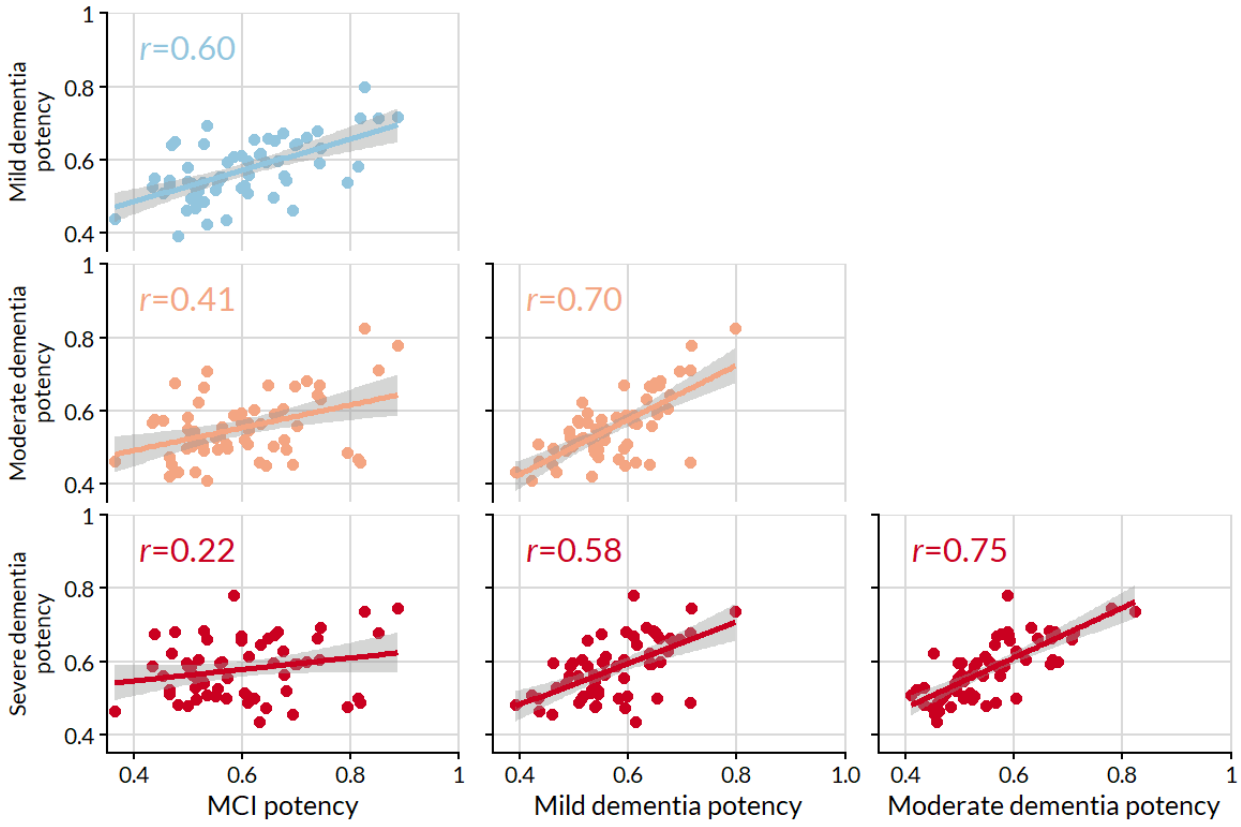


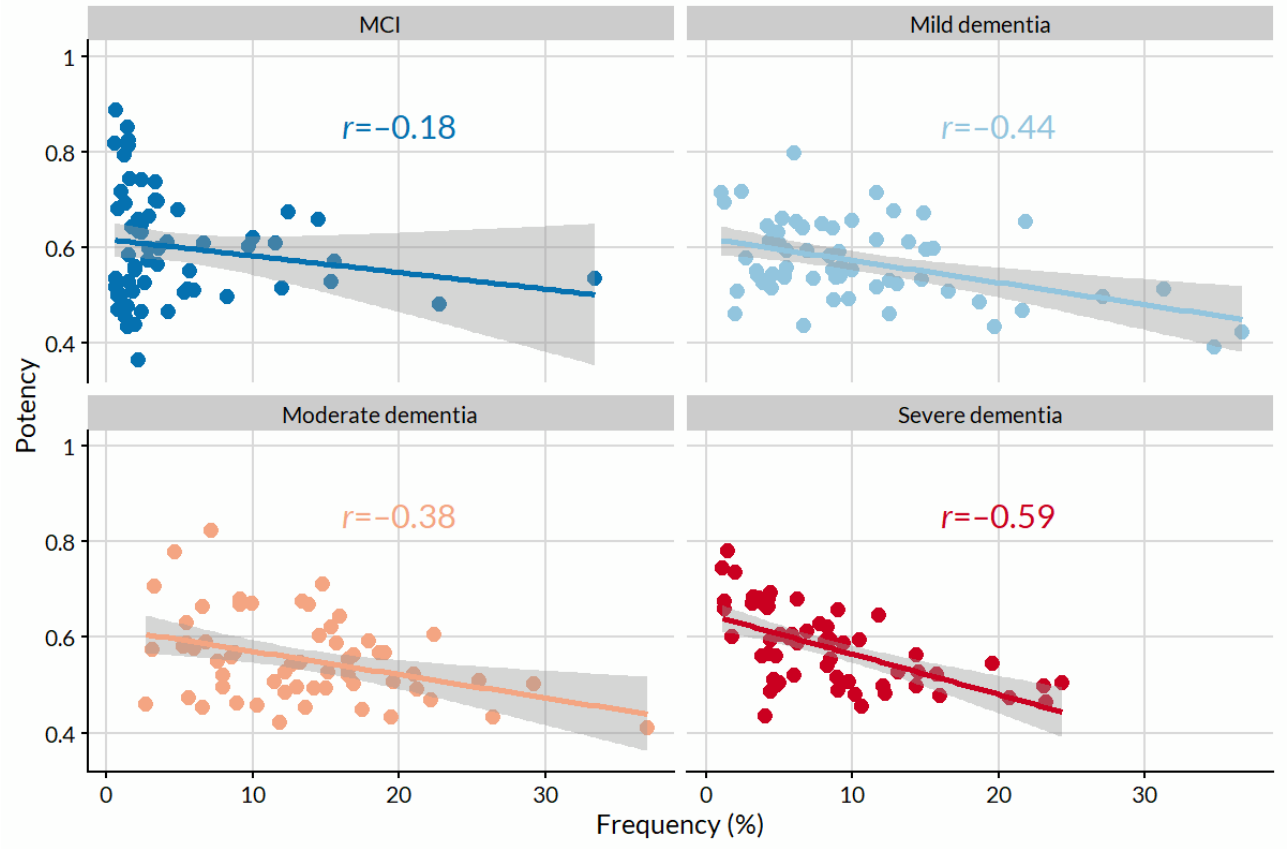
Figure 5. Pairwise relationships of symptom potency between stages. Pearson correlation coefficients (*r*) and lines of best fit (with 95% CIs) are displayed for each pair. MCI: mild cognitive impairment.



Frequency and Potency

We discovered a clear discrepancy between symptoms that were most frequent (Figure 1) and those that were most potent (Figure 3). Only interest or initiative was both highly frequent (632/2874, 21.99%) and potent (potency 0.65, 95% CI 0.63-0.67) among the mild dementia symptom profiles. Symptom frequency was negatively correlated with potency regardless of severity (Figure 6). The degree of association varied by stage from weakly correlated in MCI profiles (r=−0.18) to moderately correlated in severe dementia profiles (r=−0.59). The patterns (or trajectories) of potency and frequency are visualized for selected symptoms in Figure S1 in Multimedia Appendix 1.

Figure 6. The relationship between symptom frequency and potency, stratified by stage. Pearson correlation coefficients (r) and lines of best fit (with 95% CIs) are displayed for each stage. MCI: mild cognitive impairment.



Discussion

Principal Findings

This study used a machine learning algorithm, which was trained with clinician-staged data, to investigate how symptom tracking in web-based profiles differed by the severity of cognitive impairment (MCI and dementia). Our key finding was the large distinction between what is common and what is most important to people with cognitive impairment and their caregivers. This reinforces the importance of capturing the patients' and caregivers' voices when determining clinically meaningful changes in MCI and dementia trials. The fact that we were able to discover associations that were sensible and meaningful suggests that information collected on the web has the potential to yield useful insights into this population. As clinical treatments that focus on single-protein abnormalities understandably exclude people who do not conform to classic profiles consistent with those abnormalities, there is a need for data on real-world experiences of the larger constituency of people in whom dementia reflects a variety of disease processes.

Symptom tracking frequency was similar across MCI and mild and moderate dementia profiles, with all pairwise correlations between frequencies r>0.5. Memory of recent events and repetitive questions or stories were notably among the top 3 symptoms for all 3 stages. However, neither of these symptoms appeared among the 10 most frequent symptoms in the severe dementia profiles, which differed appreciably from the other stages (all r<0.5). With severe impairment, the early hallmarks of dementia (such as impaired memory) are present but become secondary to more distressing symptoms. In particular, we found that increasing disease severity was most associated with loss of independent function (incontinence, eating, and mobility) and behavioral problems (aggression, restlessness, and hallucinations).

The most potent (relatively important) symptoms were generally those related to disruptions in everyday life that, although less common, had great meaning when they occurred, for example, problems in looking after grandchildren, hobbies or games, and travel. Declining cognition may be a concern; however, the resulting changes to routine can be especially distressing to affected people and their families. Potency was most similar between stages of similar severity. For instance,

the strongest relationships were between the adjacent pairs of MCI–mild dementia, mild-to-moderate dementia, and moderate-to-severe dementia, with all $r \geq 0.6$.

As the large majority of our users (3792/3909, 97.01%) are caregivers, we see it reasonable to conceptualize symptom potency as an indicator of caregiver burden—the more distressing and burdensome symptoms are more likely to be ranked as highly potent by our users. In this context, we see in this study that impairment in instrumental activities of daily living (IADL; higher-order functions such as travel, meal preparation or cooking, shopping, and telephone use) place the most burden on caregivers, relative to basic activities of daily living and cognitive and behavioral symptoms. Our results are aligned with those of previous studies investigating symptoms and caregiving difficulties across the stages of the disease. A survey conducted by Alzheimer Europe [52] asked 1181 caregivers of patients with mild, moderate, or severe dementia about their current and most distressing symptoms. Problems with activities of daily living (including financial activities, shopping, cooking, and telephone use) were the most prevalent (reported by 96% of caregivers) and most problematic (68% of caregivers) symptoms, followed by behavioral symptoms (50%). In a multicenter study of 328 informal caregivers of patients who mostly experienced mild or moderate AD, IADL deficits were associated with more caregiver burden, as measured by the Zarit Burden Interview [53]. A prospective cohort study of 135 patients, ranging from those with MCI (CDR=0.5) to those with severe (CDR=3) dementia, found that depressive state in caregivers was independent of cognitive decline but was strongly associated with a decline in IADLs and delusional behavior [54].

Across stages, symptom frequency was negatively correlated with potency, and the strength of this relationship generally increased with increasing severity. We believe this pattern reflects the nature of the disease course. Early on, and especially before diagnosis, gradual change in cognitive function will be both apparent and alarming to the person living with the problems and to their caregivers; this likely underlies why the typical symptoms are still fairly potent. As deterioration increases, so does the heterogeneity in its manifestation. The typical clinical presentation becomes the accepted norm (ie, still frequent but less potent), and the impact on quality of life becomes more potent.

The contrast in symptom frequency and potency also has important implications for the measurement and interpretation of clinically meaningful changes. Outcome measures must be practical to use in that they do not overburden the informant with long interview times [55]. Including several items also has the risk of probing irrelevant information, which can affect an instrument's sensitivity to change. Striking a balance between robustness and concision is a substantial challenge in outcomes development, especially in dementia, where patient priorities are highly variable, as shown in this study. A number of best practices [56] and statistical techniques [15] exist to tackle this heterogeneity using standardized outcome measures; however, we believe that an individualized approach guided by the patient and caregiver's priorities for treatment is a simpler solution to a complex problem. By focusing on what matters most, we can guarantee that any changes measured are meaningful.

Although there are too many symptoms to compare with the literature, we draw attention to a few, such as the following: repetitive questions or stories, sleep disturbance, and interest or initiative. The stage-specific frequencies and potencies of these symptoms are visualized in Figure S1 in Multimedia Appendix 1.

We have explored the symptom of verbal repetition (here, repetitive questions or stories) in the Atlantic Canada AD Investigation of Expectations trial (open-label trial of donepezil in mild-to-moderate AD) [33] and the Video-Imaging Synthesis of Treating AD trial (randomized controlled trial of galantamine in mild-to-moderate AD) [34]. In both trials, where goal attainment scaling was the primary outcome, reduction of verbal repetition was identified as a goal of treatment in 46% and 44% of patients, respectively. This symptom notably improved more often in patients treated with galantamine than in those treated with a placebo. Our data were consistent with these secondary analyses and with a 2013 analysis of myGoalNav users [39], where verbal repetition was commonly tracked (26% overall), especially in the mild stage (35%). Hwang et al [57] also found verbal repetition to be an early sign of dementia that was troublesome to caregivers. When patients and caregivers are allowed a voice, we see that verbal repetition is important, common, and responsive, although it typically goes unmeasured by standard tests.

We found the increasing frequency of sleep disturbances with severity to be compelling, especially as it was the most commonly tracked symptom in severe dementia profiles. Similarly, Moe et al [58] found sleep disturbances to increase with disease severity in a sample of 78 AD patients. Sleep disturbances are also important at earlier stages; its prevalence in patients with MCI was estimated to be 14% to 59% in a review of 15 studies [59]. Here, the frequency was 7% in myGoalNav MCI profiles, which is unsurprisingly lower than prevalence estimates but still ranked as the 12th most frequent MCI symptom among 60 symptoms. Growing evidence suggests that sleep disturbances are a risk factor for AD [60,61], which underlines the importance of further study during predementia stages.

The symptom of interest or initiative describes a patient who is losing interest in everyday life and who has become disengaged from others and the world around them. It is common, distressing for caregivers [62], and a potential risk factor for progression from MCI to dementia [63]. In the Video-Imaging Synthesis of Treating AD trial, decreased initiation was a treatment goal for 71 of the 84 participants with mild-to-moderate AD (out of 130) who were described as having the symptom [64]. Unsurprisingly, it emerged as a noteworthy symptom in our analyses, particularly in mild dementia profiles, where it was among both the 10 most frequent and most potent symptoms. This symptom may also be important for its sensitivity to change. In a survey of caregiver and patient judgment on changes in symptoms, apathy was the neuropsychiatric symptom that improved the most in patients

with MCI and AD treated with the nutritional intervention Fortasyn Connect [65].

With no approved treatment for individuals with prodromal AD, some promise is seen in nonpharmacological interventions, especially those that combine multiple lifestyle modifications such as diet and exercise [66,67]. The Finnish Geriatric Intervention Study to Prevent Cognitive Impairment and Disability study showed that a multimodal intervention combining diet, exercise, cognitive training, and vascular risk monitoring might improve or maintain cognitive functioning among older individuals who are at risk of dementia [68]. Nutritional interventions are also being developed to tackle dietary deficiencies associated with AD pathology. A medical food (Fortasyn Connect) has been shown to improve memory in randomized controlled trials of patients with mild [69,70], but not mild-to-moderate [71], AD over 3 and 6 months of treatment. These trials were followed by the 24-month LipiDiDiet randomized controlled trial for individuals with prodromal AD [72]. There was no significant treatment effect on the neuropsychological test battery primary end point; however, there was evidence of cognitive and functional benefit, as assessed by the secondary CDR–Sum of Boxes end point, and this effect increased with better baseline cognition. In addition, results of the LipiDiDiet 36-month extension trial showed significant treatment effects on multiple measures of cognition, function, and disease progression [73]. Taken together, these studies highlight the potential for early interventions in dementia, notably with lifestyle modification, especially dietary lifestyle. With this comes a need for adequately sensitive outcomes to detect meaningful effects at early stages, for which individualized symptom tracking may be a solution [13].

Limitations

Our data must be interpreted with caution as it comprises observer-reported tracking data completed mostly by caregivers of people with dementia, who were not supervised in how they described or recorded the symptoms of the people for whom they were caring. There may have been a selection bias toward caregivers with higher functioning who can more easily locate and operate myGoalNav. Furthermore, as the myGoalNav is not a checklist of symptoms, the tracking frequencies presented here are distinct from symptom *prevalence*. There are also limitations in the development of the staging algorithm model, such as potential bias in the training data because of clinician facilitation [44].

Conclusions

Our results emphasize the importance of patient-centricity in evaluating interventions for MCI and dementia [74-76]. A personalized outcome, for example, of a grandparent being able to travel and look after their grandchild independently, will be more meaningful to the patient and their family compared with a 4-point change on the ADAS-Cog, which is considered as the main criterion for benefit. Asking patients about what is most important is sensitive to change and is inherently clinically meaningful [36,77,78]. This can be especially valuable in the predementia stages, where standard outcomes such as the ADAS-Cog and Mini Mental State Examination lack sensitivity [79,80]. Tools such as myGoalNav Dementia, and individualized outcome measures such as goal attainment scaling [81,82], are pragmatic ways of capturing the patient voice in real-world and clinical trial settings.

Acknowledgments

This study was jointly sponsored by Ardea Outcomes and Nutricia. Manuscript writing, study conceptualization, study design, data collection, and data analysis and interpretation were performed by Ardea Outcomes.

Conflicts of Interest

KR is the President, Chief Science Officer, and a shareholder of Ardea Outcomes. In the past 3 years, KR has also been a part of an advisory board for Roche and Genetech and has delivered 2 talks sponsored by Nutricia. TD, SEH, JS, AS, and SS are employees of Ardea Outcomes.

Multimedia Appendix 1
Supplementary methods detailing the logistic regression model of symptom frequency with dementia stage as a monotonic predictor. Table S1 contains all symptom frequencies by stage and Table S2 contains all symptom potencies by stage. Figure S1 depicts the patterns of potency and frequency for selected symptoms.
[DOCX File , 164 KB - jmir_v24i1e29219_app1.docx]

References

1. Dubois B, Feldman HH, Jacova C, Dekosky ST, Barberger-Gateau P, Cummings J, et al. Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria. Lancet Neurol 2007 Aug;6(8):734-746. [doi: 10.1016/S1474-4422(07)70178-3] [Medline: 17616482]
2. McKhann GM, Knopman DS, Chertkow H, Hyman BT, Jack CR, Kawas CH, et al. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):263-269 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.005] [Medline: 21514250]

3. McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984 Jul;34(7):939-944. [doi: 10.1212/wnl.34.7.939] [Medline: 6610841]
4. Jack CR, Bennett DA, Blennow K, Carrillo MC, Dunn B, Haeberlein SB, Contributors. NIA-AA research framework: toward a biological definition of Alzheimer's disease. Alzheimers Dement 2018 Apr;14(4):535-562 [FREE Full text] [doi: 10.1016/j.jalz.2018.02.018] [Medline: 29653606]
5. Boyle PA, Yu L, Leurgans SE, Wilson RS, Brookmeyer R, Schneider JA, et al. Attributable risk of Alzheimer's dementia attributed to age-related neuropathologies. Ann Neurol 2019 Jan;85(1):114-124. [doi: 10.1002/ana.25380] [Medline: 30421454]
6. Kapasi A, DeCarli C, Schneider JA. Impact of multiple pathologies on the threshold for clinically overt dementia. Acta Neuropathol 2017 Aug;134(2):171-186 [FREE Full text] [doi: 10.1007/s00401-017-1717-7] [Medline: 28488154]
7. Davis DH, Muniz Terrera G, Keage H, Rahkonen T, Oinas M, Matthews FE, et al. Delirium is a strong risk factor for dementia in the oldest-old: a population-based cohort study. Brain 2012 Sep;135(Pt 9):2809-2816 [FREE Full text] [doi: 10.1093/brain/aws190] [Medline: 22879644]
8. James BD, Wilson RS, Capuano AW, Boyle PA, Shah RC, Lamar M, et al. Hospitalization, Alzheimer's disease and related neuropathologies, and cognitive decline. Ann Neurol 2019 Dec;86(6):844-852 [FREE Full text] [doi: 10.1002/ana.25621] [Medline: 31614018]
9. Wallace LM, Theou O, Godin J, Andrew MK, Bennett DA, Rockwood K. Investigation of frailty as a moderator of the relationship between neuropathology and dementia in Alzheimer's disease: a cross-sectional analysis of data from the Rush Memory and Aging Project. Lancet Neurol 2019 Feb;18(2):177-184. [doi: 10.1016/S1474-4422(18)30371-5] [Medline: 30663607]
10. Wallace LM, Theou O, Darvesh S, Bennett DA, Buchman AS, Andrew MK, et al. Neuropathologic burden and the degree of frailty in relation to global cognition and dementia. Neurology 2020 Dec 15;95(24):e3269-e3279 [FREE Full text] [doi: 10.1212/WNL.0000000000010944] [Medline: 32989103]
11. Fratiglioni L, Marseglia A, Dekhtyar S. Ageing without dementia: can stimulating psychosocial and lifestyle experiences make a difference? Lancet Neurol 2020 Jun;19(6):533-543. [doi: 10.1016/S1474-4422(20)30039-9] [Medline: 32470425]
12. Rockwood K, Fay S, Gorman M. The ADAS-cog and clinically meaningful change in the VISTA clinical trial of galantamine for Alzheimer's disease. Int J Geriatr Psychiatry 2010 Feb;25(2):191-201. [doi: 10.1002/gps.2319] [Medline: 19548273]
13. Rockwood K, Howlett SE, Hoffman D, Schindler R, Mitnitski A. Clinical meaningfulness of Alzheimer's Disease Assessment Scale-Cognitive subscale change in relation to goal attainment in patients on cholinesterase inhibitors. Alzheimers Dement 2017 Oct;13(10):1098-1106. [doi: 10.1016/j.jalz.2017.02.005] [Medline: 28341540]
14. Sabbagh MN, Hendrix S, Harrison JE. FDA position statement "early Alzheimer's disease: developing drugs for treatment, guidance for industry". Alzheimers Dement (N Y) 2019;5:13-19 [FREE Full text] [doi: 10.1016/j.trci.2018.11.004] [Medline: 31650002]
15. Posner H, Curiel R, Edgar C, Hendrix S, Liu E, Loewenstein DA, et al. Outcomes assessment in clinical trials of Alzheimer's disease and its precursors: readying for short-term and long-term clinical trial needs. Innov Clin Neurosci 2017;14(1-2):22-29 [FREE Full text] [Medline: 28386518]
16. Alzheimer's disease: developing drugs for treatment guidance for industy. U.S. Department of Health and Human Services, Food and Drug Administration. 2018. URL: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/alzheimers-disease-developing-drugs-treatment-guidance-industy [accessed 2022-01-11]
17. Hughes CP, Berg L, Danziger WL, Coben LA, Martin RL. A new clinical scale for the staging of dementia. Br J Psychiatry 1982 Jun;140:566-572. [doi: 10.1192/bjp.140.6.566] [Medline: 7104545]
18. Morris JC. The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 1993 Nov;43(11):2412-2414. [doi: 10.1212/wnl.43.11.2412-a] [Medline: 8232972]
19. Sperling RA, Aisen PS, Beckett LA, Bennett DA, Craft S, Fagan AM, et al. Toward defining the preclinical stages of Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):280-292 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.003] [Medline: 21514248]
20. Kozauer N, Katz R. Regulatory innovation and drug development for early-stage Alzheimer's disease. N Engl J Med 2013 Mar 28;368(13):1169-1171. [doi: 10.1056/NEJMp1302513] [Medline: 23484795]
21. Brys M, Pirraglia E, Rich K, Rolstad S, Mosconi L, Switalski R, et al. Prediction and longitudinal study of CSF biomarkers in mild cognitive impairment. Neurobiol Aging 2009 May;30(5):682-690 [FREE Full text] [doi: 10.1016/j.neurobiolaging.2007.08.010] [Medline: 17889968]
22. Fox NC, Black RS, Gilman S, Rossor MN, Griffith SG, Jenkins L, AN1792(QS-21)-201 Study. Effects of Abeta immunization (AN1792) on MRI measures of cerebral volume in Alzheimer disease. Neurology 2005 May 10;64(9):1563-1572. [doi: 10.1212/01.WNL.0000159743.08996.99] [Medline: 15883317]
23. Rinne JO, Brooks DJ, Rossor MN, Fox NC, Bullock R, Klunk WE, et al. 11C-PiB PET assessment of change in fibrillar amyloid-beta load in patients with Alzheimer's disease treated with bapineuzumab: a phase 2, double-blind, placebo-controlled, ascending-dose study. Lancet Neurol 2010 Apr;9(4):363-372. [doi: 10.1016/S1474-4422(10)70043-0] [Medline: 20189881]

24. Salloway S, Sperling R, Keren R, Porsteinsson AP, van Dyck CH, Tariot PN, ELND005-AD201 Investigators. A phase 2 randomized trial of ELND005, scyllo-inositol, in mild to moderate Alzheimer disease. Neurology 2011 Sep 27;77(13):1253-1262 [FREE Full text] [doi: 10.1212/WNL.0b013e3182309fa5] [Medline: 21917766]
25. Dubois B, Villain N, Frisoni GB, Rabinovici GD, Sabbagh M, Cappa S, et al. Clinical diagnosis of Alzheimer's disease: recommendations of the International Working Group. Lancet Neurol 2021 Jun;20(6):484-496. [doi: 10.1016/S1474-4422(21)00066-1] [Medline: 33933186]
26. Bartlett SJ, Barnes T, McIvor RA. Integrating patients into meaningful real-world research. Ann Am Thorac Soc 2014 Feb;11 Suppl 2:S112-S117. [doi: 10.1513/AnnalsATS.201309-327RM] [Medline: 24559023]
27. Roadmap to patient-focused outcome measurement in clinical trials. U.S. Food and Drug Administration. 2015. URL: https://www.fda.gov/media/87004/download [accessed 2022-01-11]
28. Morel T, Cano SJ. Measuring what matters to rare disease patients - reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. Orphanet J Rare Dis 2017 Nov 02;12(1):171 [FREE Full text] [doi: 10.1186/s13023-017-0718-x] [Medline: 29096663]
29. Rockwood K, Graham JE, Fay S, ACADIE Investigators. Goal setting and attainment in Alzheimer's disease patients treated with donepezil. J Neurol Neurosurg Psychiatry 2002 Nov;73(5):500-507 [FREE Full text] [doi: 10.1136/jnnp.73.5.500] [Medline: 12397141]
30. Rockwood K, Fay S, Song X, MacKnight C, Gorman M, Video-Imaging Synthesis of Treating Alzheimer's Disease (VISTA) Investigators. Attainment of treatment goals by people with Alzheimer's disease receiving galantamine: a randomized controlled trial. CMAJ 2006 Apr 11;174(8):1099-1105 [FREE Full text] [doi: 10.1503/cmaj.051432] [Medline: 16554498]
31. Kiresuk TJ, Sherman RE. Goal attainment scaling: a general method for evaluating comprehensive community mental health programs. Community Ment Health J 1968 Dec;4(6):443-453. [doi: 10.1007/BF01530764] [Medline: 24185570]
32. Bouwens SF, van Heugten CM, Verhey FR. Review of goal attainment scaling as a useful outcome measure in psychogeriatric patients with cognitive disorders. Dement Geriatr Cogn Disord 2008;26(6):528-540 [FREE Full text] [doi: 10.1159/000178757] [Medline: 19052453]
33. Asp E, Cloutier F, Fay S, Cook C, Robertson ML, Fisk J, et al. Verbal repetition in patients with Alzheimer's disease who receive donepezil. Int J Geriatr Psychiatry 2006 May;21(5):426-431. [doi: 10.1002/gps.1486] [Medline: 16676290]
34. Rockwood K, Fay S, Jarrett P, Asp E. Effect of galantamine on verbal repetition in AD: a secondary analysis of the VISTA trial. Neurology 2007 Apr 03;68(14):1116-1121. [doi: 10.1212/01.wnl.0000258661.61577.b7] [Medline: 17404193]
35. Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, et al. The diagnosis of mild cognitive impairment due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. Alzheimers Dement 2011 May;7(3):270-279 [FREE Full text] [doi: 10.1016/j.jalz.2011.03.008] [Medline: 21514249]
36. Rockwood K. An individualized approach to tracking and treating Alzheimer's disease. Clin Pharmacol Ther 2010 Oct;88(4):446-449. [doi: 10.1038/clpt.2010.68] [Medline: 20856240]
37. Rockwood K, Zeng A, Leibman C, Mucha L, Mitnitski A. Validation of an informant-reported web-based data collection to assess dementia symptoms. J Med Internet Res 2012 Mar 12;14(2):e42 [FREE Full text] [doi: 10.2196/jmir.1941] [Medline: 22411293]
38. Rockwood K, Mitnitski A, Richard M, Kurth M, Kesslak P, Abushakra S. Neuropsychiatric symptom clusters targeted for treatment at earlier versus later stages of dementia. Int J Geriatr Psychiatry 2015 Apr;30(4):357-367. [doi: 10.1002/gps.4136] [Medline: 24798635]
39. Reeve E, Molin P, Hui A, Rockwood K. Exploration of verbal repetition in people with dementia using an online symptom-tracking tool. Int Psychogeriatr 2017 Jun;29(6):959-966 [FREE Full text] [doi: 10.1017/S1041610216002180] [Medline: 28274302]
40. McGarrigle L, Howlett SE, Wong H, Stanley J, Rockwood K. Characterizing the symptom of misplacing objects in people with dementia: findings from an online tracking tool. Int Psychogeriatr 2019 Nov;31(11):1635-1641. [doi: 10.1017/S104161021800220X] [Medline: 30698122]
41. Rockwood K, Stanley J, Dunn T, Howlett S. PIH2 development and validation of a patient centered outcome measure for use in dementia drug trials. Value Health 2019 May;22:S183. [doi: 10.1016/j.jval.2019.04.803]
42. Rockwood K, Richard M, Leibman C, Mucha L, Mitnitski A. Staging dementia from symptom profiles on a care partner website. J Med Internet Res 2013 Aug 07;15(8):e145 [FREE Full text] [doi: 10.2196/jmir.2461] [Medline: 23924608]
43. The Delphi Method: Techniques and Applications. Boston: Addison-Wesley; 1975.
44. Shehzad A, Rockwood K, Stanley J, Dunn T, Howlett SE. Use of patient-reported symptoms from an online symptom tracking tool for dementia severity staging: development and validation of a machine learning approach. J Med Internet Res 2020 Nov 11;22(11):e20840 [FREE Full text] [doi: 10.2196/20840] [Medline: 33174853]
45. Rockwood JK, Richard M, Garden K, Hominick K, Mitnitski A, Rockwood K. Precipitating and predisposing events and symptoms for admission to assisted living or nursing home care. Can Geriatr J 2014 Mar;17(1):16-21 [FREE Full text] [doi: 10.5770/cgj.17.93] [Medline: 24596590]
46. Bürkner PC, Charpentier E. Modelling monotonic effects of ordinal predictors in Bayesian regression models. Br J Math Stat Psychol 2020 Nov;73(3):420-451. [doi: 10.1111/bmsp.12195] [Medline: 31943157]

47. Bürkner P. brms: an R Package for Bayesian multilevel models using Stan. J Stat Softw 2017;80(1):1-28 [FREE Full text] [doi: 10.18637/jss.v080.i01]
48. Bürkner P. Advanced Bayesian multilevel modeling with the R Package brms. R J 2018;10(1):395-411 [FREE Full text] [doi: 10.32614/rj-2018-017]
49. Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, et al. Stan: a probabilistic programming language. J Stat Softw 2017;76(1):1-32 [FREE Full text] [doi: 10.18637/jss.v076.i01]
50. R: a language and environment for statistical computing. R Foundation for Statistical Computing. 2018. URL: https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing [accessed 2022-01-11]
51. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the Tidyverse. J Open Sci Softw 2019 Nov;4(43):1686. [doi: 10.21105/joss.01686]
52. Georges J, Jansen S, Jackson J, Meyrieux A, Sadowska A, Selmes M. Alzheimer's disease in real life--the dementia carer's survey. Int J Geriatr Psychiatry 2008 May;23(5):546-551. [doi: 10.1002/gps.1984] [Medline: 18232054]
53. Tsai C, Hwang W, Lee J, Wang W, Huang L, Huang L, et al. Predictors of caregiver burden in aged caregivers of demented older patients. BMC Geriatr 2021 Jan 14;21(1):59 [FREE Full text] [doi: 10.1186/s12877-021-02007-1] [Medline: 33446114]
54. Hasegawa N, Hashimoto M, Koyama A, Ishikawa T, Yatabe Y, Honda K, et al. Patient-related factors associated with depressive state in caregivers of patients with dementia at home. J Am Med Dir Assoc 2014 May;15(5):371.e15-371.e18. [doi: 10.1016/j.jamda.2014.02.007] [Medline: 24685406]
55. Sheehan B. Assessment scales in dementia. Ther Adv Neurol Disord 2012 Nov;5(6):349-358 [FREE Full text] [doi: 10.1177/1756285612455733] [Medline: 23139705]
56. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL. Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 2018;6:149 [FREE Full text] [doi: 10.3389/fpubh.2018.00149] [Medline: 29942800]
57. Hwang JP, Tsai SJ, Yang CH, Liu KM, Lirng JF. Repetitive phenomena in dementia. Int J Psychiatry Med 2000;30(2):165-171. [doi: 10.2190/2QDA-YAL3-2E69-PYJW] [Medline: 11001279]
58. Moe KE, Vitiello MV, Larsen LH, Prinz PN. Symposium: cognitive processes and sleep disturbances: sleep/wake patterns in Alzheimer's disease: relationships with cognition and function. J Sleep Res 1995 Mar;4(1):15-20 [FREE Full text] [doi: 10.1111/j.1365-2869.1995.tb00145.x] [Medline: 10607136]
59. Beaulieu-Bonneau S, Hudon C. Sleep disturbances in older adults with mild cognitive impairment. Int Psychogeriatr 2009 Aug;21(4):654-666. [doi: 10.1017/S1041610209009120] [Medline: 19426575]
60. Shi L, Chen S, Ma M, Bao Y, Han Y, Wang Y, et al. Sleep disturbances increase the risk of dementia: a systematic review and meta-analysis. Sleep Med Rev 2018 Aug;40:4-16. [doi: 10.1016/j.smrv.2017.06.010] [Medline: 28890168]
61. Minakawa EN, Wada K, Nagai Y. Sleep disturbance as a potential modifiable risk factor for Alzheimer's disease. Int J Mol Sci 2019 Feb 13;20(4):803 [FREE Full text] [doi: 10.3390/ijms20040803] [Medline: 30781802]
62. Bianchetti A, Trabucchi M. Behavioural and psychological symptoms of dementia: clinical aspects. Neurosci Res Commun 2005 Nov;35(3):173-183. [doi: 10.1002/nrc.20031]
63. Vicini Chilovi B, Conti M, Zanetti M, Mazzù I, Rozzini L, Padovani A. Differential impact of apathy and depression in the development of dementia in mild cognitive impairment patients. Dement Geriatr Cogn Disord 2009;27(4):390-398. [doi: 10.1159/000210045] [Medline: 19339777]
64. Cook C, Fay S, Rockwood K. Decreased initiation of usual activities in people with mild-to-moderate Alzheimer's disease: a descriptive analysis from the VISTA clinical trial. Int Psychogeriatr 2008 Oct;20(5):952-963. [doi: 10.1017/S1041610208007230] [Medline: 18405399]
65. Bianchetti A, Perotta D, Cravello L, Ranieri P, Trabucchi M. Effectiveness of a specific nutritional supplement on cognitive, behavioral and functional symptoms in mild cognitive impairment and Alzheimer's dementia: caregivers judgments. Results of an observational survey. J Gerontol Geriatrics 2018;66(2):74 [FREE Full text]
66. Kivipelto M, Mangialasche F, Ngandu T. Lifestyle interventions to prevent cognitive impairment, dementia and Alzheimer disease. Nat Rev Neurol 2018 Nov;14(11):653-666. [doi: 10.1038/s41582-018-0070-3] [Medline: 30291317]
67. Livingston G, Sommerlad A, Orgeta V, Costafreda SG, Huntley J, Ames D, et al. Dementia prevention, intervention, and care. Lancet 2017 Dec 16;390(10113):2673-2734. [doi: 10.1016/S0140-6736(17)31363-6] [Medline: 28735855]
68. Ngandu T, Lehtisalo J, Solomon A, Levälahti E, Ahtiluoto S, Antikainen R, et al. A 2 year multidomain intervention of diet, exercise, cognitive training, and vascular risk monitoring versus control to prevent cognitive decline in at-risk elderly people (FINGER): a randomised controlled trial. Lancet 2015 Jun 06;385(9984):2255-2263. [doi: 10.1016/S0140-6736(15)60461-5] [Medline: 25771249]
69. Scheltens P, Kamphuis PJ, Verhey FR, Olde Rikkert MG, Wurtman RJ, Wilkinson D, et al. Efficacy of a medical food in mild Alzheimer's disease: a randomized, controlled trial. Alzheimers Dement 2010 Jan;6(1):1-10.e1. [doi: 10.1016/j.jalz.2009.10.003] [Medline: 20129316]
70. Scheltens P, Twisk JW, Blesa R, Scarpini E, von Arnim CA, Bongers A, et al. Efficacy of Souvenaid in mild Alzheimer's disease: results from a randomized, controlled trial. J Alzheimers Dis 2012;31(1):225-236. [doi: 10.3233/JAD-2012-121189] [Medline: 22766770]

71. Shah RC, Kamphuis PJ, Leurgans S, Swinkels SH, Sadowsky CH, Bongers A, et al. The S-Connect study: results from a randomized, controlled trial of Souvenaid in mild-to-moderate Alzheimer's disease. Alzheimers Res Ther 2013;5(6):59 [FREE Full text] [doi: 10.1186/alzrt224] [Medline: 24280255]
72. Soininen H, Solomon A, Visser PJ, Hendrix SB, Blennow K, Kivipelto M, LipiDiDiet clinical study group. 24-month intervention with a specific multinutrient in people with prodromal Alzheimer's disease (LipiDiDiet): a randomised, double-blind, controlled trial. Lancet Neurol 2017 Dec;16(12):965-975 [FREE Full text] [doi: 10.1016/S1474-4422(17)30332-0] [Medline: 29097166]
73. Soininen H, Solomon A, Visser PJ, Hendrix SB, Blennow K, Kivipelto M, LipiDiDiet clinical study group. 36-month LipiDiDiet multinutrient clinical trial in prodromal Alzheimer's disease. Alzheimers Dement 2021 Jan;17(1):29-40 [FREE Full text] [doi: 10.1002/alz.12172] [Medline: 32920957]
74. Reuben DB, Tinetti ME. Goal-oriented patient care--an alternative health outcomes paradigm. N Engl J Med 2012 Mar 01;366(9):777-779. [doi: 10.1056/NEJMp1113631] [Medline: 22375966]
75. Frank LB. Commentary on "Health economics and the value of therapy in Alzheimer's disease." Report from the Alzheimer's Association Research Roundtable on patient-reported outcomes and dementia research. Alzheimers Dement 2007 Jul;3(3):162-165. [doi: 10.1016/j.jalz.2007.04.383] [Medline: 19595932]
76. Forsythe L, Heckert A, Margolis MK, Schrandt S, Frank L. Methods and impact of engagement in research, from theory to practice and back again: early findings from the Patient-Centered Outcomes Research Institute. Qual Life Res 2018 Jan;27(1):17-31 [FREE Full text] [doi: 10.1007/s11136-017-1581-x] [Medline: 28500572]
77. Rockwood K, MacKnight C. Assessing the clinical importance of statistically significant improvement in anti-dementia drug trials. Neuroepidemiology 2001 May;20(2):51-56. [doi: 10.1159/000054761] [Medline: 11359070]
78. Shabbir SH, Sanders AE. Clinical significance in dementia research: a review of the literature. Am J Alzheimers Dis Other Demen 2014 Sep;29(6):492-497 [FREE Full text] [doi: 10.1177/1533317514522539] [Medline: 24526758]
79. Spering CC, Hobson V, Lucas JA, Menon CV, Hall JR, O'Bryant SE. Diagnostic accuracy of the MMSE in detecting probable and possible Alzheimer's disease in ethnically diverse highly educated individuals: an analysis of the NACC database. J Gerontol A Biol Sci Med Sci 2012 Aug;67(8):890-896 [FREE Full text] [doi: 10.1093/gerona/gls006] [Medline: 22396476]
80. Raghavan N, Samtani MN, Farnum M, Yang E, Novak G, Grundman M, Alzheimer's Disease Neuroimaging Initiative. The ADAS-Cog revisited: novel composite scales based on ADAS-Cog to improve efficiency in MCI and early AD trials. Alzheimers Dement 2013 Feb;9(1 Suppl):S21-S31 [FREE Full text] [doi: 10.1016/j.jalz.2012.05.2187] [Medline: 23127469]
81. Jennings LA, Ramirez KD, Hays RD, Wenger NS, Reuben DB. Personalized goal attainment in dementia care: measuring what persons with dementia and their caregivers want. J Am Geriatr Soc 2018 Nov;66(11):2120-2127 [FREE Full text] [doi: 10.1111/jgs.15541] [Medline: 30298901]
82. Jennings LA, Palimaru A, Corona MG, Cagigas XE, Ramirez KD, Zhao T, et al. Patient and caregiver goals for dementia care. Qual Life Res 2017 Mar;26(3):685-693 [FREE Full text] [doi: 10.1007/s11136-016-1471-7] [Medline: 28000094]

Abbreviations

AD: Alzheimer disease
ADAS-Cog: Alzheimer Disease Assessment Scale–Cognitive subscale
CDR: Clinical Dementia Rating
IADL: instrumental activities of daily living
MCI: mild cognitive impairment

Edited by A Mavragani; submitted 08.04.21; peer-reviewed by L McGarrigle, C Kerssens, P Harvey; comments to author 28.06.21; revised version received 13.08.21; accepted 02.12.21; published 27.01.22.

Please cite as:
Dunn T, Howlett SE, Stanojevic S, Shehzad A, Stanley J, Rockwood K
Patterns of Symptom Tracking by Caregivers and Patients With Dementia and Mild Cognitive Impairment: Cross-sectional Study
J Med Internet Res 2022;24(1):e29219
URL: https://www.jmir.org/2022/1/e29219
doi:10.2196/29219
PMID:35084341

Publisher:
JMIR Publications
130 Queens Quay East.
Toronto, ON, M5A 3Y5
Phone: (+1) 416-583-2040
Email: support@jmir.org